Individual Differences as Affected by Practice

Gates

INDIVIDUAL DIFFERENCES AS AFFECTED BY PRACTICE

BY
GEORGINA STICKLAND GATES, Ph. D.

ARCHIVES OF PSYCHOLOGY
EDITED BY R. S. WOODWORTH
No. 58

NEW YORK
AUGUST, 1922

AGENTS: G. E. STECHERT & CO. : London (2 Star Yard, Carey St., W. C.); Paris (16, rue de Condé)

ARCHIVES OF PSYCHOLOGY

SUB-STATION 84, NEW YORK CITY. The Subscription price is five dollars per volume of about 500 pages. Volume I comprises Nos. 1-10, Volume II, Nos. 11-18, Volume III, Nos. 19-25, Volume IV, Nos. 26-32, Volume V, Nos, 33-39, Volume VI, Nos. 40-46, Vol. VII, Nos. 47-52, and Vol. VIII, Nos. 53-58. The available numbers are as follows:

2. On the Functions of the Cerebrum: The Frontal Lobes: SHEPHERD IVORY FRANZ. 50c.
3. Empirical Studies in the Theory of Measurements: EDWARD L. THORNDIKE. 50c.
4. Rhythm as a Distinguishing Characteristic of Prose Style: ABRAM LIPSKY. 50c.
5. The Field of Distinct Vision: W. C. RUEDIGER. 70c.
6. The Influence of Bodily Position on Mental Activities: ELMER E. JONES. 50c.
7. A Statistical Study of Literary Merit: FREDERIC LYMAN WELLS. 30c.
8. The Relation between the Magnitude of the Stimulus and the Time of Reaction: SVEN FROEBERG. 35c.
9. The Perceptual Factors in Reading: FRANCIS MARION HAMILTON. 50c.
10. Time in English Verse Rhythm: WARNER BROWN. 70c.
11. The Hearing of Primitive Peoples: FRANK G. BRUNER. $1.00.
12. Studies in Development and Learning: EDWIN A. KIRKPATRICK. $1.00
13. The Inaccuracy of Movement: H. L. HOLLINGWORTH. 80c.
14. A Quantitative Study of Rhythm: HERBERT WOODROW. 60c.
16. On certain Electrical Processes in the Human Body and their Relation to Emotional Reactions: FREDERIC LYMAN WELLS and ALEXANDER FORBES. 40c.
17. The Relative Merit of Advertisements: EDWARD K. STRONG, JR. $1.00.
18. Attention and Movement in Reaction Time: J. V. BREITWIESER. 50c. (Cloth, 75c.)
19. An Empirical Study of Certain Tests for Individual Differences: MARY THEODORA WHITLEY. $1.25. (Cloth, $1.50.)
20. Visual Acuity with Lights of Different Colors and Intensities: DAVID EDGAR RICE. 50c. (Cloth, 75c.)
21. The Curve of Forgetting: C. H. BEAN. 45c. (Cloth, 70c.)
22. The Influence of Caffein on Mental and Motor Efficiency: H. L. HOLLINGWORTH. $1.50. (Cloth, $1.75.)
23. Reaction Time to Retinal Stimulation. A. T. POFFENBERGER, JR. 70c. (Cloth, 95c.)
24. Interference and Adaptability: ARTHUR JEROME CULLER. 75c. (Cloth, $1.00.)
25. Reaction to Multiple Stimuli: JOHN WELHOFF TODD. 60c. (Cloth, 85c.)
26. A Study in Incidental Memory: GARRY C. MYERS. $1.00. (Cloth, $1.25.)
27. A Statistical Study of Eminent Women: CORA SUTTON CASTLE. 80c. (Cloth, $1.05.)
28. The Mental Capacity of the American Negro: MARION J. MAYO. 60c. (Cloth, 85c.)
29. Experimental Studies in Judgment: H. L. HOLLINGWORTH. $1.25. (Cloth, $1.50.)
30. The Psychological Researches of James McKeen Cattell: A Review by Some of His Pupils. $1.00. (Cloth, $1.25.)
31. Fatigue and Its Effects upon Control: ISAAC EMERY ASH. 60c. (Cloth, 85c.)
32. The Transfer Effects of Practice in Cancellation Tests: MELVIN ALBERT MARTIN. 60c. (Cloth, 85c.)
33. The Intellectual Status of Children Who are Public Charges: J. L. STENQUIST, E. L. THORNDIKE, and M. R. TRABUE. 50c. (Cloth, 75c.)
34. The Relation of Quickness of Learning to Retentiveness: DARWIN OLIVER LYON. 50c. (Cloth, 75c.)
35. The Overcoming of Distraction and Other Resistances: JOHN J. B. MORGAN. 75c. (Cloth, $1.00.)
36. The Psychology of the Negro—An Experimental Study: GEORGE OSCAR FERGUSON, JR. $1.25. (Cloth, $1.50.)
37. The Effect of Distraction on Reaction Time: JOHN E. EVANS: $1.00. (Cloth, $1.25.)
38. The Effect of Humidity on Nervousness and on General Efficiency: LORLE IDA STECHER. 90c. (Cloth, $1.15.)
39. The Mechanism of Controlled Association: MARK A. MAY. 75c. (Cloth, $1.00.)
40. Recitation as a Factor in Memorizing: ARTHUR I. GATES. $1.00. (Cloth, $1.25.)
41. Mental Fatigue during Continuous Exercise of a Single Function: THOMAS RUSSELL GARTH. 85c. (Cloth, $1.10.)
42. A Psychological Study of Trade-Mark Infringement: RICHARD H. PAYNTER, JR. 85c. (Cloth, $1.10.)
43. Individual Differences and Family Resemblances in Animal Behavior: HALSEY J. BAGG. 70c. (Cloth, $1.00.)
44. Experimental Studies in Recall and Recognition: EDITH MULHALL ACHILLES. 90c. (Cloth, $1.25.)
45. The Morphologic Aspect of Intelligence: SANTE NACCARATI. 70c. (Cloth, $1.10.)
46. Psychological Examinations of College Students: F. EDITH CAROTHERS. $1.25.
47. The Effects of Practice on Judgments of Absolute Pitch: EVELYN GOUGH. $1.25
48. An Experimental Study of Silent Thinking: RUTH S. CLARK. $1.40.
49. Some Empirical Tests in Vocational Selection: HERBERT W. ROGERS. 75 cents.
50. Adenoids and Diseased Tonsils: Their Effect on General Intelligence. MARGARET CORB. $1.00.
51. An Experimental Study of the Factors and Types of Voluntary Choice. ALFRED H. MARTIN. $1.50.
52. Some Well-Known Mental Tests Evaluated and Compared. DOROTHY RUTH MORGENTHAU. 80 cents.
53. Mood in Relation to Performance. ELIZABETH T. SULLIVAN. $1.00.
54. The Influence of Incentive and Punishment upon Reaction Time: ALBERT M. JOHANSON. 80c.
55. Psychological Tests Applied to Factory Workers: EMILY THORP BURR. $1.25.
56. A Study of the Relation of Accuracy to Speed: HENRY E. GARRETT. $1.25.
57. An Experimental Study of Hunger in Its Relation to Activity: TOMI WADA. $1.50.
58. Individual Differences as Affected by Practice: GEORGINA STICKLAND GATES. $1.00.
59. Studies in Industrial Psychology: ELSIE OSCHRIN BREGMAN. 75 cents.

INDIVIDUAL DIFFERENCES AS AFFECTED BY PRACTICE

BY
GEORGINA STICKLAND GATES, Ph. D.

ARCHIVES OF PSYCHOLOGY
EDITED BY R. S. WOODWORTH
No. 58

NEW YORK
AUGUST, 1922

AGENTS: G. E. STECHERT & CO. : London (2 Star Yard, Carey St., W. C.); Paris (16, rue de Condé)

TABLE OF CONTENTS

CHAPTER		PAGE
1.	Problems and Methods of Investigation	5
2.	Changes in Individual Scores and Group Relations with Practice	10
3.	Inter-correlations of the Tests	19
4.	The Relation of Improvement to Initial Ability	37
5.	The Relation of Improvement to Final Ability	63
6.	The Relation of Improvement in one Test to Improvement in Another and of Initial and Final Gain	68
7.	Summary	72

ACKNOWLEDGMENT

I am indebted to Professor Thorndike for a number of suggestions as to statistical procedure; to Professor Woodworth for advice concerning changes to be made in the presentation of the study; and to Professor H. L. Hollingworth for the original problem, for the opportunity to conduct the experiments, for the use of valuable records, and for constant interest and encouragement.

This paper was completed, aside from some revision of the text, before October 1, 1919.

Individual Differences as Affected by Practice

CHAPTER ONE

PROBLEMS AND METHODS OF INVESTIGATION

In the literature of psychology, a large number of studies have been included under the term "practice experiments." Investigations in this field range from studies of learning Russian or telegraphy to inquiries concerning improvement in ability to tap. Experiments have been performed upon children; upon adults under normal conditions, as when working in a psychological laboratory; upon adults in various unusual conditions as when dosed with caffein or alcohol, or working in a humid atmosphere; observations have been made of the effect of the repetition of a situation upon monkeys, dogs, cats, chickens, turtles, paramocium. Such problems as the determination of the factors in learning, of the typical course of improvement, of the adequate representation of learning, the necessity for and causes of plateaus, and the question of the probable limit of improvement and of the frequency of approach to such a limit, have been attacked many times both by theoretical analyses, and by the method of "patience, starving out, and harrassing to death."

The majority of such investigations, whether mainly theoretical or mainly practical in their implications, are concerned with changes in individual scores. Attention is directed to the subject's initial and final status and to variations observed in his measured efficiency. On the basis of such objective records, an attempt is made to determine the factors responsible for improvement, retrogression or for apparent lack of change. By varying conditions, the influence of external and internal events upon the effectiveness of the repetition, as determined by the score, are studied.

Such investigations deal mainly with phenomena in which individuals are assumed to resemble each other. As we assume that all children of a given physical age will vary about a norm, just as do a certain random sampling—say 500—children of that age, so it may be suggested that most people's

learning of telegraphy will resemble, at least in important factors, the amount, rate, and kind of improvement found in the five persons studied. Discoveries concerning the effect of practice have been, in the main, results obtained from careful experimentation on a few persons. In typical experiments, for example, Romer, Weygant, Vogt, Kapemen, and Swift, had each one subject; Rejall, Hill and Book had two; Bair had three; Coover and Angell four; Bryan, Harter, Swift, Heiman, and James had five; Thorndike, Starch and Fracker had eight; Whitley had nine; Wells, Hahn and Dearborn had ten; Kline, Leuba and Hyde had seventeen; Munn had twenty-three. We have a few studies like those of Kirby who had seven hundred; Brown and Peterson with thirty-four; Thorndike with forty-three. For the most part the number of subjects has been small, and in those cases where a large number of individuals were studied, the functions tested were usually but one or two.

The interest in the results drawn from the individual scores of a few persons studied intensively was in the improvement phenomena itself, rather than in any further problem. The authors inquire what is learning; under what conditions does it come about; what factors are important for improvement; how may we facilitate it, rather than what effect does the knowledge gained from practice experiments have upon some other psychological concept.

The present study differs from the classical investigations in that it is not concerned primarily with individual scores but with relations between individual records, in that individual differences rather than individual similarities are stressed, and in that the practice phenomena are studied only with a view to the light which they may throw upon a further question—that of the nature of native or acquired differences between the subjects of the investigation.

For this reason, we inquire whether the relative position of individuals in the group remains constant or changes with practice; whether there is any relation between proficiency in the various functions before and after opportunity for improvement; we attempt to discover the factors which determine initial and final status; and the relation between gains made in different tests or in different parts of the practice period.

The main questions in the order in which they are discussed are:—

1. What changes in individual scores occur as a result of the repetition of a performance?
2. Does the variability of scores decrease with continued practice?
3. Do group relations remain unchanged after repetition? Does the best person remain at the top and the worst person at the bottom? Is the first trial diagnostic of the succeeding trials and how well does the last trial correlate with those preceding?
4. What is the correlation between the functions? Does this change with practice?
5. What is the correlation between initial ability and improvement?
6. How are final capacity and amount of gain correlated?
7. Is gain in one test indicative of gain in another?
8. Is gain at the beginning of the practice period diagnostic of gain toward the end?

Method

The following table gives the tests used and the method of scoring.

TESTS AND METHOD OF SCORING

Color-Naming: Two Woodworth-Wells blanks pasted together. Record, time required to name correctly a set of two hundred colors. Test repeated twenty-five times.

Tapping: Record, number of taps executed in two minutes with the hand stylus, right hand. Test repeated twenty-five times.

Adding: Kraepelin blank. Record, time required to add one hundred examples of two numbers each. Test repeated thirty times.

Multiplying: Record, time required to solve mentally a set of five multiplication examples. The examples required the multiplying of one two-place by another two-place number. No number below four was used. Different examples were used each time up to the twentieth trial. The twenty-first and twenty-second sets were repetitions of the first and second sets of examples.

Word-building: Record, number of shorter words built from a given word (as weather, psychiatry, etc.) in a specified time. Words varied in difficulty and time allowed was changed as the experiment proceeded. Twenty different words used. The

twenty-first and twenty-second words were the same as the first and second words.

The tests, it will be noted, vary from the very simple tapping performance in which we may assume that practice means the exercise of relatively few bonds, to the more complex tests such as multiplying and word-building in which the subject's reaction, performed each time in response to a different problem, is the result of a comparatively elaborate cerebral process. The tests differ also in their content, and in the method of scoring. One record is concerned with a motor performance; two have to do with words, and two with numbers. Three functions are scored by the time and three by the amount method. An equal number of trials of color-naming, adding and multiplying means, then, an equal number of trials successfully performed regardless of time spent. Practice trials of tapping and word-building, on the other hand, refer to definite temporal periods.

The subjects were twenty-three women students taking their second or third terms work in psychology. The group, though not more homogeneous than groups used for similar experiments, were much more alike in age, experience, interest, and intelligence, than twenty-three people chosen at random.

All the tests, except word-building were given to the subjects individually, by one experimenter. Each subject came to the experimenter for two separate half-hours each week. The time of day, exact interval between test periods varied, of course, for different subjects, but for any one individual, the time and regularity of performance were kept as constant as possible.

For the final results, twenty-nine trials of the adding test and twenty-two trials of each of the other tests were used. All computations are based on a score which is either the first trial, the very best trial, the median of three, or eleven, or (in adding) four trials, or the best of three or four trials. For the correlations, the formula $\rho = 1 - \frac{6\,\Sigma d^2}{n(n^2-1)}$ was used. ρ has not been changed into r.

A table of probable errors of coefficients computed for twenty-five individuals, and one showing the divisions of the prac-

tice period which will be used in chapters two and three, follow:

*Table of Probable Errors**

Using Formula: $P.E. = \frac{.67\ (1 - r^2)}{n}$

r	Group n = 23
.1	.14
.2	.14
.3	.13
.4	.12
.5	.10
.6	.09
.7	.07
.8	.05
.9	.03

The Divisions of the Practice Series

1st Division. The initial trial.

2nd Division. The second, third, and fourth trial of color-naming, tapping, multiplying, word-building. The second, third, fourth and fifth trial of adding.

3rd Division. Fifth, sixth, seventh of four tests. Sixth, seventh, eighth, ninth of adding.

4th Division. Eighth, ninth, tenth of four tests. Tenth, eleventh, twelfth, thirteenth of adding.

5th Division. Eleventh, twelfth, thirteenth of four tests. Fourteenth, fifteenth, sixteenth, seventeenth of adding.

6th Division. Fourteenth, fifteenth, sixteenth of four tests. Eighteenth, nineteenth, twentieth, twenty-first of adding.

7th Division. Seventeenth, eighteenth, nineteenth of four tests. Twenty-second, twenty-third, twenty-fourth, twenty-fifth of adding.

8th Division. Twentieth, twenty-first, twenty-second of four tests. Twenty-sixth, twenty-seventh, twenty-eighth, twenty-ninth of adding.

*This formula, it should be noted, is strictly applicable only to the Pearson coefficients.

CHAPTER TWO

CHANGES IN INDIVIDUAL SCORES AND GROUP RELATIONS WITH PRACTICE

Before we can ask what changes relations between individuals undergo as a result of practice we must know exactly what effect repetition in the present case has upon individual scores. At the risk of appearing to prove the obvious, we must assure ourselves of the existence, amount and relative rate of improvement in these tests.

We may ask what transformation takes place in the average of the group as a whole. Does the average time required decrease or increase? Is the average amount of work performed larger or smaller? This information, though valuable, is not sufficient. It does not tell us enough about individual improvement or retrogression. A small amount of gain in the average of the group might mean a large improvement on the part of some persons and a stationary position or retrogression by others. In order to tell the story adequately, it might seem necessary to present all the records in full, but as such tables would be cumbersome and, as all the correlations in Chapter Three and Four are based upon median or best scores, be unnecessary for a correct interpretation of the facts. Here is included a table (Table 1) showing the group average and the individual scores for the first trial and the group average and individual median for the other seven divisions (see previous Chapter) of four of the tests.

Because of the varying difficulty of the words used and the change in time allowed for word-building, it is difficult to obtain from this test information as to amount of improvement, change in rate, or approximation to a practice level. For this reason it is omitted here.

If we consider first group averages—found in the last line of Table 1—we see that there is gradual improvement in all tests. The average of the adding improves up to the sixth division, color-naming to the seventh division, multiplying and tapping improve throughout the practice period. The improvement in tapping and color-naming is relatively small, in multiplying and adding there is a relatively large gross

amount of improvement. In tapping the improvement is approximately one-twelfth of the final score, in color-naming it is one-fifth, in adding one-half and in multiplying about twice as great as the final record.

The gross gain as measured is greatest at the beginning of the practice series.* From the first to the third periods there is a change in averages of 12.6 seconds in color-naming, 45.3 taps in tapping, 30.2 seconds in adding, 232 seconds in multiplying. Between the sixth and eighth periods there is a change of only 8 seconds in color-naming, of 7. 4 taps in tapping, a gain of 8 seconds in adding, a drop of 59 seconds in multiplying.

Adjacent differences are given in the following table:

TABLE NO. 2

Adjacent Differences in Average Scores.

	Δ1-2	Δ2-3	Δ3-4	Δ4-5	Δ5-6	Δ6-7	Δ7-8
Color-Naming	9.7	2.9	.5	3.1	.8	1.6	8
Tapping	18.7	26.6	–.7	11.3	6.3	5.6	1.8
Adding	21.1	9.1	2.3	.5	2.8	.5	.3
Mult.	214	18	15	29.2	23.7	12.5	46.1

The largest difference is (with the exception of tapping) between the initial trial and the median of the next three. All but multiplying show a very small or entire lack of improvement at the end.†

Inspection of individual scores reveals the same phenomenon —a relatively large amount of improvement at the beginning which becomes less as practice continues. At the end of the period there may be little or no change in score. The particular question of individual progression will be discussed in Chapters Four and Five. The individual scores, presented here, we need only observe bear out, in general, the results noted from a comparison of average cases.

The changes observed in individual records may or may not produce a change in the relative position of the subjects. A second question is, consequently,—Do persons improve uniformly with reference to one another? Does the best person tend to stay at the top, the worst person at the foot, the me-

*Our method of scoring may, of course, be inaccurate. It may be that gains made at the end of the series should be given a higher score. The improvement curve may really be a straight line. This is not questioned here. Reference is made only to gross measured scores.

†The larger improvement noted in multiplying may have been due, not only to the fact that the function was a more complex one with more opportunity for improvement, but to the repetition at the end of the practice series of the first two multiplication examples. All subjects denied the slightest memory of these examples. Yet the large amount of gain noted there may be due to some subliminal familiarity.

diocre individual in the middle range? Is a record made in an initial test at all indicative of what ability would be at the end of twenty-two trials? Correlating the first trial with the medians of the other seven divisions, we obtain the following table. In this it is possible to include word-building, for changes in difficulty of words or in time spent, as they are constant at each trial for all persons, do not affect conclusions based upon relative position.

TABLE NO. 3

Correlation—First Trial with Median of Other Seven Divisions in Same Test.

	2	3	4	5	6	7	8
C. N.	78	74	85	75	77	79	72
Tap.	62	39	53	45	53	29	37
Add.	45	36	45	48	44	49	33
Mul.	42	45	53	57	50	33	67
W. B.	82	85	83	75	70	81	72
Average	62	56	64	60	59	54	56
Ave., excluding multiplying	67	59	67	61	61	60	54

There is a shift in relative positions as a result of practice. But increasing the distance between the compared trial does not decrease intercorrelations as much as one would expect. The median of the second, third and fourth trials resembles more closely (with the exception of multiplying)* the initial record than does the median of the twentieth, twenty-first and twenty-second repetitions. But the difference is slight, with the exception of tapping and possibly word-building—well within the probable error of the average coefficient. Averages in fact are practically identical. There is a suggestion that correlations do decrease but the difference in figures is too small and the appearance of high coefficients too variable for any proof.

A correlation of the median records in the second division with the median records of the other six divisions gives similar results. There, coefficients are all, as we would expect from their increased reliability, a little higher than those in Table No. 3. The correlations given by tapping and multiplying decrease. The others are practically unchanged.

When the median of the last division is correlated with the medians of the other divisions the coefficients found are, in general, higher than those obtained in Tables Three and Four. This may be explained as due to the increased reliability of the last trial.

*The final high correlation of multiplying may be due in part to the repetition at the end of the series, of the first two sets of examples. This does not seem to affect word-building.

TABLE NO. 4

Median of Second Division Correlated with Median of Other Six Divisions.

	3rd	4th	5th	6th	7th	8th
Color-naming	90	89	88	92	87	90
Tapping	72	79	78	71	54	63
Adding	85	89	89	86	74	82
Multiplying	64	50	55	35	19	43
Word-building	79	85	75	68	73	77
Average	78	78	77	70	61	71

TABLE NO. 5

Correlation of Last Division with Other Divisions.

	1st	2nd	3rd	4th	5th	6th	7th
C. N.	72	90	89	85	91	94	94
Tap.	37	63	92	95	92	94	86
Add.	33	82	70	80	84	84	86
Mul.	67	43	68	82	80	80	74
W. B.	72	77	80	79	75	80	86
Ave.	56	71	80	84	84	86	85

The coefficients of tapping, adding and word-building increase. Multiplying is irregular and color-naming, after the first division is uniformally high. The average does not increase after the fourth division. A correlation of .85 or more is reached in the color-naming test in the second division, in tapping in the third, in adding and word-building in the eighth, in multiplying in the second. Evidenty four trials of a color-naming test will give a fairly reliable index of an individual's final capacity as displayed after 22 trials of the test. Seven are needed for tapping, twenty-two for adding and word-building. The correlation of multiplying varies so much that it is unreliable.

Similar correlations have been found by Whitley, Wells, Chapman, Brown and Hollingworth.* The decrease in Hollingworth's result is marked. The small size of many of the coefficients and the changes observed in the course of the experiment are attributable, we believe, to the great length of the practice period. Individuals, in this experiment, did actually reach their limit of improvement. The table is repeated on page 14.

In the present experiment it seems as though individuals maintained their positions fairly well, though there is a definite amount of change. In some cases the difference increases as the distance between repetitions is increased. This appears most plainly in Table No. 5 where the most reliable

*Hollingworth "Individual Differences Before, During and After Practice."

TABLE NO. 6

Correlation of Order of Position of Thirteen Individuals Before, During and After Practice.

The correlation is in each case with the final order, after 175 practice trials (in two cases 130 trials).

Test	Preliminary Trial	5th Trial	25th Trial	50th Trial	30th Trial	130th Trial	175th Trial
Adding	154	193	874	869	973	962	1,000
Opposites	088	616	490	835	945	984	1,000
Colors	682	891	858	913	968	968	1,000
Discrimination	676	621	604	500	500	785	1,000
Cancellation	665	676	885	686	934	1,000	....
Co-ordination	528	793	770	902	946	1,000	....
Tapping	231	484	627	682	693	885	1,000
	41	61	73	79	85	92	1,000

score (final record) is used as the basis for comparison. We might attribute lack of higher correlation between initial scores to unreliability of the determination. Correction for attenuation would probably raise initial coefficients more than final so the difference, presumably, would appear more plainly. On the other hand, as the figures stand, in many cases correlations between first and last records vary only as they might by chance from correlations between first and second division scores.

Beside observing changes in individual scores and in position as compared with initial rank, it is necessary to note the relative variability at the beginning and end of the practice period. How reliable are the records which we have selected as measures of individual ability? Is there any change in reliability of scores with increased practice? We have determined the reliability of the first trial, of the median of the second or third division, of the median of the fourth or sixth division, and of the median of the seventh or eighth division. The reliability of the initial trial is obtained by correlating it with the median of the next three trials. The reliability of the medians of the second or third division is obtained from the adjacent correlation between the two. The correlation between the fifth and sixth, and seventh and eighth medians likewise gives their reliability.

TABLE NO. 7

Self-Correlation of Median Records.

Test	Cor. 1 & 2	2 & 3	4 & 5	7 & 8
Color-naming	78	90	95	94
Tapping	62	72	91	86
Adding	45	85	95	86
Multiplying	42	64	68	74
Word-building	82	79	75	86
Average	62	78	85	85

The median of the last three trials has on the average a higher correlation of adjacent medians than has the median of the second and third trials, which is similarly more reliable than the initial trial alone. In the simpler functions, trials near but not at the end of the practice period (divisions 5 & 6) are the most reliable. Considering the first trial only the order of reliability of the tests seems to be word-building, color-naming, tapping, adding and mutiplying. An observation of the last trial shows that color-naming is the least variable, word-building, adding and tapping come next and multiplying, which shows throughout a lower self-correlation, last.

Another measure which will be employed for certain computations, is one which we have called "best" records. Instead of taking the median of three measures in a practice division as a measure of individual ability as observed at that time, we have taken the highest of these three records. Such a score might serve to eliminate various sources of error (as distraction, temporary lack of interest, fatigue) which would operate more frequently to lower than to raise individual records. On the other hand a "best" score, since it is a determination based upon only one record, might be raised above its actual amount by some chance factor, an error made in recording or some "accident in the nervous system." If such a measure is to be employed, the reliability of the scores must be determined and a comparison of the variability of median and of best records is valuable. Self-correlations of best records at the beginning and at the end of practice obtained in the same way as were self-correlations of median records give us the following table.

TABLE NO. 8

	2 & 3	7 & 8
C. N.	90	96
Tap.	72	94
Add.	81	89
Mul.	59	64
W. B.	73	71
Average	75	83

With the exception of word-building the correlations increase. The final figures for color-naming, tapping and adding are slightly higher than corresponding figures for median records, multiplying and word-building have each a lower correlation in best records as compared with median scores. On the whole, the best record is fairly reliable. There is scarcely more or less variability than there is in the median

record. The more complex functions show a slightly greater regularity when measured by median records, the simpler functions seem to vary at the last trial less when scores are computed in terms of best records.

The facts here suggested are in agreement, in general, with those found by other students of practice. We have found an improvement after repetition, larger in some functons than others, greater in all cases at the beginning than at the end of the period of repetition. This decrease in improvement of individual scores is paralleled by a decrease in amount of change made in the relative position of the individuals. Where scores are highest (about the sixth or seventh periods) reliability is greatest. Where the largest amount of improvement through practice is taking place—as at the beginning of the series—there is the largest amount of variability in relative position. First scores in some cases resemble less and less closely other records, as the number of intervening repetitions is increased. Practice in these tests, then, makes for an increase in scores. With this increase there goes a more or less slight shift in relative position, which is greatest where individual scores and group averages show the largest amount of gain.

TABLE NO. 1

Median Record for Each Individual in Each Division of the Practice Series.

Color-Naming.

Divisions Individual.	1	2	3	4	5	6	7	8
A	117	114	113	120	110	106	107	101
B	098	88	90	89	91	91	85	84
C	203	137	146	149	143	132	138	140
D	122	123	110	120	114	117	107	111
E	113	115	109	97	107	111	103	105
F	100	101	101	99	101	96	99	108
G	105	95	103	100	97	94	95	96
H	110	108	95	96	94	97	87	98
I	075	73	73	72	68	70	66	69
J	123	93	86	98	88	78	80	80
K	110	107	100	98	95	94	88	91
L	096	90	98	90	92	89	83	89
M	095	96	97	92	91	85	82	83
N	124	122	118	108	108	105	104	107
O	115	93	97	96	93	93	96	92
P	097	94	88	95	87	86	87	91
Q	103	89	84	91	88	89	82	82
R	110	97	100	95	92	92	89	93
S	117	117	106	115	98	99	97	102
T	133	122	110	112	99	108	110	108
U	132	114	110	112	101	106	107	106
V	111	105	98	103	102	96	59	60
W	084	76	70	67				
Av.	112.7	103	100.1	100.6	96.5	95.7	94.1	94.9

TABLE NO. 1
Tapping.

Divisions Individual	1	2	3	4	5	6	7	8
A	686	748	706	747	743	748	764	792
B	720	805	766	797	803	794	825	801
C	703	716	635	680	635	668	689	708
D	699	828	806	748	795	718	785	832
E	745	654	654	639	610	610	604	645
F	845	886	826	838	167	862	880	896
G	863	895	943	1001	1104	1087	1040	1051
H	850	882	950	965	918	926	870	904
I	824	825	983	905	879	895	815	930
J	910	888	912	855	927	904	831	882
K	769	740	762	840	756	789	756	769
L	843	748	788	792	795	811	824	803
M	755	787	794	807	801	912	848	811
N	665	645	790	639	686	685	709	668
O	770	791	737	810	925	934	931	919
P	735	749	780	741	767	735	726	727
Q	714	670	684	710	758	807	870	834
R	615	710	824	846	845	860	898	931
S	728	822	820	855	840	860	850	788
T	725	695	764	737	716	675	722	700
U	713	730	787	790	761	734	870	834
V	686	686	675	675	690	695	704	715
W	692	795	919	872	1028	976	912	924
Average	750.6	769.3	795.9	795.2	806.5	812.8	818.4	820.2

TABLE NO. 1
Adding.

Divisions Individual	1	2	3	4	5	6	7	8
A	103	97.5	89.0	82.5	83.5	83.5	84.0	80.0
B	124	87.0	76.5	70.5	72.5	64.4	66.5	69.5
C	118	94.5	92.0	87.5	89.0	84.0	81.5	82.0
D	094	83.5	78.5	75.5	73.0	73.0	73.5	73.5
E	090	85.5	75.5	74.0	70.5	69.0	69.0	75.5
F	097	98.5	89.5	92.0	90.5	84.0	70.5	80.0
G	129	88.5	76.5	76.5	73.5	70.5	71.0	72.0
H	094	88.0	78.0	77.0	78.0	74.0	72.0	79.0
I	101	89.5	75.0	78.0	77.0	71.5	76.5	73.0
J	077	82.5	71.5	65.0	65.0	64.0	63.0	64.5
K	132	98.5	85.0	83.0	80.5	77.5	76.0	79.5
L	984	70.5	63.5	61.5	56.5	56.0	55.5	55.5
M	096	89.5	79.0	74.5	75.0	69.0	72.5	74.0
N	102	76.5	68.5	67.1	66.5	64.0	67.0	64.5
O	094	74.5	69.5	67.0	61.5	63.0	65.5	60.5
P	122	79.1	78.1	78.1	75.6	76.3	82.3	81.6
Q	101	66.1	63.5	64.0	67.1	62.4	62.0	61.7
R	103	75.1	68.6	62.9	62.4	61.2	65.3	63.4
S	115	85.0	74.6	70.9	73.9	70.3	71.2	65.9
T	113	78.0	64.5	62.0	64.3	63.1	62.4	59.5
U	142	94.0	77.5	80.1	80.5	75.3	76.0	76.3
V	112	88.0	77.0	73.3	71.8	71.3	73.0	71.5
W	82	69.0	59.1	54.9	56.1	53.8	56.8	57.1
Average	105.4	84.3	75.2	72.9	72.4	69.6	70.1	70.4

TABLE NO. 1
Multiplying.

Divisions Individual	1	2	3	4	5	6	7	8
A	303	538	370	282	223	200	215	152
B	665	485	320	310	255	270	231	199
C	1104	360	393	447	376	374	358	321
D	730	268	217	239	203	290	172	177
E	613	474	203	182	266	199	169	151
F	877	311	405	735	439	360	524	424
G	480	316	376	238	272	263	263	165
H	374	371	406	259	230	180	231	166
I	778	707	504	554	539	522	429	373
J	541	405	334	198	244	214	227	246
K	1011	786	545	409	308	270	229	228
L	317	196	179	124	136	137	155	139
M	373	337	259	374	375	355	264	216
N	408	294	284	332	210	277	297	174
O	425	420	290	292	193	236	145	165
P	822	440	405	385	488	264	224	258
Q	642	364	302	217	224	194	162	129
R	606	232	205	209	187	183	205	166
S	423	357	440	322	295	310	351	190
T	282	224	285	120	230	89	122	75
U	457	525	423	479	462	379	280	262
V	420	345	208	245	177	224	184	105
W	133	111	87	159	103	101	167	53
Average	556	342	324	309	279.8	256.1	243.6	197.1

CHAPTER THREE

INTER-CORRELATIONS OF THE TESTS

In the course of the experimental investigation of general intelligence, a number of theories concerning the correlation of mental abilities have been advanced. Two opposing hypotheses may be defined—the compensation theory and the theory of the positive correlation of desirable traits. Or the problem may be observed from a slightly different angle, and one may distinguish three hypotheses, the theory of specific habits, of types or levels, and of the common factor.

The theory of compensation has, of course, been abandoned by psychologists, though it is still retained in such popular beliefs as that slow learning means long retention, or that ability at abstract thinking will incapacitate one for housework or gardening. Inverse relations are not, however, unknown. Such instances as Chapman's[1] negative correlation between cancellation and multiplication, and McCall's[2] negative correlation between cancellation and a number of other mental tests, may be advanced to show, at the will of the critic, either the probability of the existence of such inverse relations, or the unreliability of the experiments. In the majority of cases positive correlation seems to occur.

The hypothesis of the specification of traits has, like the compensation theory, been abandoned. It is no longer held that each mental function is a separate entity which may be developed independently, without effect upon other abilities.

Concerning the cause of such positive correlation as is observed, the two theories, the doctrine of levels and of the common factor are in dispute today. The difference seems to be main one of emphasis. Spearman,[3] who is the exponent of the "common factor," writes: "Here the view supported is that all performances depend to a certain degree upon one and the same general common factor, provisionally termed 'Gen-

[1]Chapman, J. C.—Individual Differences in Ability and Improvement and their Correlations. C. U. Contrib. to Ed.

[2]McCall.—Correlations of some Psychological and Educational Measurements—T. C. Contrib. to Ed.

[3]Hart and Spearman.—General Ability, Its Existence and Nature B. J. P. Volume 5.

eral Ability'. Correlations are thus produced between all sorts of performances, the amount of correlation being simply proportional to the extent that the performances concerned involve the use of this general common factor or 'General Ability.' The common factor is more closely related to each function than any trait is to any other trait." Spearman admits, however, special relating forces which, though subordinate in importance to the common factor, may also make for positive correlation.

Thorndike[2] emphasizes these special relating forces rather than the common factor. He says, speaking of the possibility of discovering the relation between traits. "Other things being equal two functions would be correlated in proportion as efficiency in them depends upon the status of the unknown 'Common Factor' variously called General Intellect, Mental Energy, Ability to Learn, and the like. Perhaps if we could define this factor, learn its symptoms and test its strength in different individuals, we should find that it alone explained the main features of all cross lines (representing relations.) Perhaps on the other hand, the other relating force would be even more important." One of the special relating forces is the organization of neurones into sensory and associative levels. "Correlations seem to be closer within the analytical or abstracting functions than between these and others. So also within the purely mental associative functions like adding, completing words, giving opposites and naming objects, than between one of them and one of the sensori-motor functions. The sensations seem to interrelate only loosely; and any one of them would relate very loosely to the associative or analytic functions, even when the latter was busied with data from that sense." Besides this organization into levels, other relating forces might make for positive correlation. Examples of these would be instinctive tendencies—as interest in persons might produce high correlations between love of ceremonies, ability in sociology and interest in literature—and the influence of special training—as the correlation between knowledge of Latin and of Geometry might be caused by the training of many individuals in both lines.

Almost every psychologist who has given more than two or three tests to a number of individuals, has made some contribution to the psychology of correlation. Those investigations

[2]Thorndike.—Educational Psychology, Volume III.

which have been most fruitful from the point of view of general theory have been attempts in the main to answer one or all of these three questions: 1st—Is there a positive correlation of desirable traits? 2nd—If such a positive correlation exists, is it due largely to the presence of a common factor or to special relating forces? 3rd—If a common factor does exist, what is its nature? Three kinds of experimental procedure have been adopted for the solution of these problems. The first method is that of giving a number of tests to a homogenous group of subjects and correlating the results of the tests with each other and with some estimate of intelligence. Ability in the second group of investigations is judged mainly by status as determined by achievement in life. There a notably good and a notably poor group are studied, and the main object of the investigation is to determine the difference in reaction of the two. The third type of study is not concerned with estimates of intelligence or with previously demonstrated ability, but simply with the relation between the traits measured by the tests themselves. Such investigations have frequently been undertaken for some other purpose—as to measure the effect of humidity or of caffein upon efficiency. Only as a secondary problem has the question of correlation been attacked.

Among the notable experiments of the first type are those of Burt, Wyatt, and Brown. Here the main interest seems to be in the definition of intelligence.

Burt* gave 12 tests to 43 boys. The tests measured sensory discrimination, motor ability, sensori-motor functions, association and voluntary attention. Teachers and students estimated the intelligence of the 43 boys. Burt correlated these estimates of intelligence with the records made by the boys in the various tests. The test of voluntary attention showed the highest correlation, association and memory tests came next, then sensori-motor, then motor, and finally tests of sensory discrimination showed the lowest correlations of all. He concludes from his results that all the functions of the human mind, the simplest and most complicated alike are processes within a single system. A process typical of higher psycho-physical levels may however, be connected with a process typical of lower psychophysical levels far less intimately than either is with a process of intermediate levels.

*Burt—Experimental Tests of General Intelligence—Br. Jr. Ps. 3.

In this Burt seems to favor the theory of levels but his emphasis—like Spearman's—is upon the common factor rather than upon the special relating forces. Voluntary attention, especially the power of readjustment to relatively novel situations is, for him, the essential factor in this general intelligence.

Wyatt[2] gave 15 tests to 75 children, giving each test twice. Analogies and completion tests, which seemed to measure functions of the higher levels, correlated most closely with estimated intelligence. Memory tests showed the next highest correlations. Wyatt, too, emphasizes the common factor.

Brown[3] who examined 259 children with 12 tests—each one given twice—found that first the completion test and next the mechanical memory tests, correlated most closely with general intelligence. Brown, unlike the other two, found no evidence of the existence of the common factor.

The second type of investigation includes such studies as those of Norsworthy, Terman, Binet and Simpson. Norsworthy,[4] in comparing defective with normal children, found that the order in which the different abilities tested would correlate with intelligence was: 1st—abstraction and association; then, memory, then various forms of perception; and last, motor control. Terman[5] tested seven bright and seven dull boys and found the bright boys superior in all the mental tests as puzzles, word-building, solving problems. Superiority of the dull boys in motor tests was explained as due to greater maturity, or different interests. Binet,[6] in testing five intelligent and six unintelligent pupils, found that tests of accuracy in tactile sensibility, counting rhythmic sounds, copying figures, sentences and drawings, memory of figures and cancellation of letters best differentiated the bright boys from the dull. He believed that quickness of adaptation was the important factor in intelligence.

Simpson,[7] made an intensive study of two groups of persons—a good group made up of 17 professors and advanced students of Columbia and a poor group of 20 men who had

[2]Wyatt—The Quantitative Investigation of Higher Mental Processes. B. J. P. 6.

[3]Brown—Some Experimental Results in the Correlation of Mental Abilities, B. J. P. 3.

[4]Norsworthy—The Psychology of Mentally Deficient Children.

[5]Terman—Genius and Stupidity—Pd. S. C. 13.

[6]Binet—"Attention et Adaptation" Annee Psychol. 6,248.

[7]Simpson—"Correlations of Mental Abilities" T. C. Contrib. to Ed.

never held any position demanding a high grade of intelligence, some from the Salvation Army Industrial Home, some from a Mission on the Bowery. He finds no hierachial arrangement of abilities, but does find a close relation among certain mental abilities which would imply a something which might be called "general intelligence." Certain capacities are, however, relatively specialized. The abilities which correlate most closely with other traits as shown by the fifteen tests given are—first, selective thinking, then memory and association, quickness and accuracy of perception, motor control, sensory discrimination. General intelligence, then, implies these different abilities in the order named.

The foregoing results (with the exception of Binet's) and a number of similar ones are based upon the scores of initial tests, upon records corrected for attenuation, or scores attained after a preliminary fore-exercise. Correlations vary from such coefficients as 94 (raw coefficients) obtained by Simpson between memory of words and Ebbinghaus test (when both groups were combined) to such numbers as 62 obtained by Burt between sorting and sound discrimination, and 25 obtained by Brown between accuracy in addition and the Ebbinghaus test. Are such figures a valid measure of the relation between the traits studied? Obviously they are not. The first trial is a poor measure of ability. Correction for attenuation is a statistical method for eliminating the influence of variability—it needs, however, confirmation from actual data. The method of giving a few practice trials does away with the influence of "adaptation." Various other conditioning factors (as previous experience of some individuals) are not excluded by this method.

Sufficient practice in the capacities to be observed might eliminate these sources of error. A score which was the result of a number of testings would be more reliable than a score based upon one or two tests. Taken at the close of the practice period, it would be subject to less variability. Since the amount of practice to be added is the same for each individual, the increment would serve to equalize the opportunity of each. In such an investigation no hypotheses concerning the nature of general intelligence or the cause of correlations observed to exist need be advanced. Such a study need throw light upon the problem of the existence of the common factor, only in so far as these further experiments give a

basis for questioning or confirming former results. When we ask what influence a further factor—as practice in a number of tests—has upon their intercorrelations, we are inquiring whether coefficients already found are reliable and representative figures.

Five practice experiments may be reported. A number of other studies as those of Whitley, Wells, Brown, Chapman, Binet, and Thorndike (see next chapter) bear only indirectly on the problem.

Spearman and Krueger[1] found that correlations for 8 successive quarter hours between such abilities as cancellation, adding, counting, reading, learning by heart tended to increase in size as practice increased, up to a certain point, after which they tended to diminish. The decrease might be due to fatigue entering as a disturbing factor. Where practice means improvement in individual scores correlations seem to increase. After fatigue enters, they diminish.

Burt[2] found that the second trial of eleven tests correlated less closely with estimated intelligence than did the first trial. As Jones points out this may be due to some factor of adaptation, common to all tests, which makes intercorrelations between tests at the first trial higher than would otherwise be the case.

Abelson[3] tested subnormal children and found the correlations between tests to be 32 (initial trial), 36, 37, and 40 at the second, third, and fourth trials.

Winch[4] studied the effect of practice on correlation between a simple motor test of cancelling all letters, and a complex motor test, cancelling a. n, o, s. The correlations on six successive days were 29, 44, 59, 48, 50, 47—an observable increase up to the third day.

Jones[5] in a study of eight tests made upon 203 boys, for four successive years, finds "no marked tendency for correlations of test measurements either to increase or to decrease on account of age and experience, under the conditions of the ex-

[1]Spearman and Krueger—Die Korrelation zwischen verschiedenen geistigen Leistungsfahigkeiten Z. f. P. 44.

[2]Burt—"Experimental tests of General Intelligence"—B. J. P. 3.

[3]Abelson—"The Measurement of Mental Ability of Backward Children, B. J. P. 4.

[4]Winch—Reported in Brown—"The Effect of Observational Errors" B. J. P. 6.

[5]Jones—"Influence of Age and Experience on Correlations."

periment. There are many striking irregularities which are difficult to explain in the course of the correlation from one year to the next."

"On the whole," he says, "it is evident that many tests decrease their correlations from the first to the second year, due presumably to the factor of the understanding of instructions (common to all tests at the time of their first administration). After this initial drop in the amount of correlation between the tests, there is likely to be a slight increase, in the following years, in the amount of correation. Whether this is due to practice alone or whether to the factors of age and vocational experience varying widely among the subjects, it is impossible to say. We are inclined to believe that the vocational life in Cincinnati during these three years does aid in differentiating the good from the bad. This is in spite of the fact that, in the case of certain types of tests, there is an evident evening up process (as in the memory tests) so that those who are proficient in one are not correspondingly proficient in the other, even after practice. The increase in the amount of correlation between different tests is greater when the tests are apparently unlike each other than when the tests are alike."

The results of Burt and Abelson are based each on only three or four trials. Spearman and Kreuger's results are complicated by the entrance of fatigue, Jones' correlations are influenced not only by practice but by the entire experience of the subjects during the year that intervened between testings. Hollingworth's* results are obtained from a very large number of trials. His experiment was such that it was possible to isolate the effect of practice from fatigue, growth, or general experience.

The following table was obtained by averaging the intercorrelations between six tests given to 13 individuals 205 times. After the initial trial, each score used is the average of five records.

TABLE NO. 9

Intercorrelations Between Tests in Hollingworth's Experiment.

	Adding	Opposites	Color Naming	Discrimination	Co-ordination	Tapping	Final Average
1	19	10	15	-07	-15	17	065
5	41	26	15	35	21	32	280
25	50	35	43	27	03	35	320
80	55	43	53	31	18	34	390
205	48	62	61	35	34	52	490

Correlations obviously increase. The change from first to second is, as Jones has pointed out, due partly to the fact that the coefficient for the first trial is based upon one score, for the second period upon an average of five scores. A decrease in variability—as these coefficients are uncorrected—may, similarly, be a factor in the increase. Hollingworth* suggests as causes in addition to this change in variability a change in the nature of the tests themselves, and possibly some explanation in terms of a common fund of energy or at least of the existence, when final ability is approximated, of a positive relation between desirable traits.

It is the object of the present chapter to investigate the effect of practice somewhat more minutely, to confirm or question the former results, and to suggest, if possible, some cause for whatever relation is found to exist.

In order to eliminate the change in the nature of the tests, a factor which seems to be of some importance in Hollingworth's results, functions were chosen which, it is believed, did not become more similar as repetitions were increased. Hollingworth's opposites and calculation test, for example, came, after identical repetition to resemble greatly the color-naming test. No conscious choosing of an opposite or mathematical operation was involved. Each stimulus on the card was bound to one definite response. With the exception of the larger number of stimulus-response series involved, and possible inequality of previous practice, these tests gradually approached in character the color-naming test. Two of the tests of the present investigation, as was noted above, cannot become automatic through repetition. The multiplying and word-building tests involve a response each time, to a new situation. Nor do these functions come to resemble each other or the color-naming, tapping, and adding tests as time goes on. Only the adding and color-naming tests might seem to approach each other. Part of the correlation between color-naming and adding might possibly be due to increased similarity. But the amount of initial likeness is not great—not nearly so important as in the tests in the former investigation—and the amount of practice given is probably insufficient for a reduction of the adding operation to a practically automatic response.

*Hollingworth—"Correlation of Abilities as affected by Practice." Jr. Ed. Psy. 4.

The practice series was divided into eight periods and each test was correlated with every other test at each of the eight divisions. Table 10 gives the result of this correlation. There the median record (in all but the first trial) is the score used.

TABLE NO. 10

Intercorrelations Between Tests (Median the Measure). Average along side without multiplying. Average at bottom with multiplying.

	C. N	Tap.	Add.	Mul.	W. B.	Av.
			1st Division			
C. N.	x	32	20	–02	21	24
Tap.	32	x	24	–16	12	23
Add.	20	24	x	29	24	23
Mul.	–02	–16	29	x	21	
W. B.	21	12	24	20	x	19
Av.	19	13	24	08	19	
			2nd Division			
C. N.	x	32	32	–12	21	28
Tap.	32	x	–21	–02	20	10
Add.	32	–21	x	51	42	18
Mul.	–12	–02	51	x	32	
W. B.	21	20	42	32	x	29
Av.	18	07	26	17	28	
			3rd Division			
C. N.	x	36	35	03	08	26
Tap.	36	x	24	–16	05	22
Add.	35	24	x	48	27	29
Mul.	03	–16	48	x	26	x
W. B.	08	05	27	26	x	13
Av.	21	12	34	15	17	
			4th Division			
C. N.	x	33	41	17	35	36
Tap.	33	x	–01	–05	25	19
Add.	41	–01	x	78	33	24
Mul.	17	–05	78	x	17	
W. B.	35	25	33	17	x	31
Av.	32	13	38	27	28	
			5th Division			
C. N.	x	63	35	–04	33	44
Tap.	63	x	26	10	40	43
Add.	35	26	x	76	46	36
Mul.	–04	10	76	x	07	
W. B.	33	40	46	07	x	40
Av.	32	35	46	22	32	
			6th Division			
C. N.	x	67	44	11	22	44
Tap.	67	x	25	05	30	40
Add.	44	25	x	56	33	34
Mul.	11	05	56	x	28	
W. B.	22	29	33	28	x	28
Av.	36	32	40	25	28	
			7th Division			
C. N.	x	42	37	16	41	40
Tap.	42	x	36	–02	03	27
Add.	37	36	x	47	20	31
Mul.	16	–02	47	x	13	
W. B.	41	03	20	13	x	21
Av.	34	20	35	19	19	

8th Division

C. N.	x	46	36	08	16	33
Tap.	46	x	22	09	–07	20
Add.	36	22	x	62	39	32
Mul.	08	09	62	x	17	
W. B.	16	–07	39	17	x	16
Av.	27	18	40	24	16	

Table 11 gives the average correlation of each test with every other test at each of the eight divisions.

TABLE NO. 11

Average Correlation of Each Test with Every Other Test.

	1	2	3	4	5	6	7	8
C. N.	19	18	21	32	32	36	34	27
Tap.	12	01	12	13	35	32	20	18
Add.	24	26	34	38	46	40	35	40
Mul.	08	17	15	27	22	25	19	24
W. B.	19	28	17	28	32	28	19	16
Av.	16	19	20	28	33	32	26	25
Ave. of Two.			24		33		26	

There seems to be a gradual increase in the average of all coefficients which lasts up to the sixth period, after which there is a decrease. The averages for the separate tests show this. Color-naming and multiplying increase up to the sixth, tapping, adding and word-building up to the fifth. An observation of the original coefficients in Table 4 confirms this. The correlation between tapping and adding reaches its maximum at the seventh division, between tapping and color-naming, color-naming and adding, word-building and multiplying at the sixth, between tapping and multiplying, tapping and word-building, word-building and adding at the fifth, between color-naming and multiplying, mutiplying and adding at the fourth, between word-building and multiplying at the second. After the maximum point there is a slight decrease in correlations.

The multiplying test has a low and occasionally negative correlation with all tests except adding. The unreliability of the test as compared with others (see chapter 3) may account for some of this observed lack of relation. Another table from which all correlations with multiplying were omited in averaging is given here.

With the exception of adding (and occasionally of word-building) all average coefficients are higher when multiplying is omitted. These correlations increase as did the others, up to the fifth or sixth division, after which there is again a slight decrease.

TABLE NO. 12

Average Coefficients—Multiplying Omitted.

Test	1	2	3	4	5	6	7	8
C. N.	24	28	26	36	44	44	40	33
Tap.	23	10	22	19	43	40	27	20
Add.	23	18	29	24	36	34	31	32
W. B.	19	29	13	31	40	28	21	16
Ave.	22	21	23	28	41	37	30	25
Average of Two			26		39		28	

Beside this median record another measure of individual achievement was employed. Best scores instead of median records in each practice period were correlated. Such scores are (See chapter 3) no more variable than median records, and this in spite of the fact that they are determinations based upon only one testing. They would seem at first sight to have the advantage of being most representative of ultimate achievement. A tendency to inaccuracy would likewise be immediately suggested, though such a suggestion seems to be contradicted by the relatively high adjacent correlation. Is there any difference in the size of correlations obtained by this method? What factors operate to produce such relations as are observed?

Table 13 gives the result of inter-correlations of best records in the eight periods.

TABLE NO. 13

Intercorrelations Between Tests (Best Record Used as Measure). Average without multiplying at side, with multiplying at bottom.

	C. N.	Tap.	Add.	Mul.	W. B.	Av.
			2nd Division			
C. N.	x	31	34	–05	15	27
Tap.	31	x	–10	01	22	14..
Add.	34	–10	x	46	37	20
Mul.	–05	01	46	x	16	
W. B.	15	22	37	16	x	25
Av.	19	11	27	15	23	
			3rd Division			
C. N.	x	28	50	38	35	38
Tap.	28	x	06	–20	28	21
Add.	50	06	x	67	11	22
Mul.	38	–20	67	x	–70	
W. B.	35	28	11	–70	x	25
Av.	38	11	34	04	01	
			4th Division			
C. N.	x	43	48	13	23	38
Tap.	43	x	07	–01	17	22
Add.	48	07	x	60	40	32
Mul.	13	–01	60	x	05	
W. B.	23	17	40	05	x	27
Av.	32	17	39	19	21	

			5th Division			
C. N.	x	57	45	04	43	48
Tap.	57	x	19	06	37	38
Add.	45	19	x	65	53	39
Mul.	04	05	65	x	03	
W. B.	43	37	53	03	x	44
Av.	37	30	46	20	34	
			6th Division			
C. N.	x	57	47	−14	30	45
Tap.	57	x	33	−19	20	37
Add.	47	33	x	58	35	38
Mul.	−14	−19	58	x	21	
W. B.	30	20	35	21	x	28
Av.	30	23	43	12	27	
			7th Division			
C. N.	x	62	42	−04	34	46
Tap.	62	x	25	−23	−04	28
Add.	42	25	x	18	28	32
Mul.	−03	−23	19	x	07	
W. B.	34	−04	28	07	x	19
Av.	34	15	29	−01	16	
			8th Division			
C. N.	x	50	44	14	30	41
Tap.	50	x	26	−20	39	38
Add.	44	26	x	52	36	35
Mul.	14	−20	52	x	−01	
W. B.	30	39	36	−01	x	35
Av.	35	24	40	11	26	

In Table 14 the correlations of each test with every other test are averaged. In Table 15 the averages are given without multiplying.

TABLE NO. 14

Average Correlation of Each Test with Every Other Test (Best Record used as Measure).

Test	2	3	4	5	6	7	8
C. N.	19	38	32	37	30	34	35
Tap.	11	11	17	30	23	15	24
Add.	27	34	39	46	43	29	40
Mul.	15	04	19	20	12	−01	11
W. B.	23	01	21	34	27	16	26
Av.	19	18	28	33	27	19	27
Av. of Two		23		30		23	

TABLE NO. 15

Average Correlation of Each Test, with Every Other Test, excluding Multiplying (Best Record used as Measure).

Test	2	3	4	5	6	7	8
C. N.	27	38	38	48	45	46	41
Tap.	14	21	22	38	37	28	38
Add.	20	22	32	39	38	32	35
W. B.	25	25	27	44	29	19	35
Av.	22	27	30	42	37	31	37
Av. of Two		29		40		34	

A comparison of these tables with those of median records shows them to be similar, though some correlations by the best record method are higher, some lower. The average coefficients are a little lower by the best record method when multiplying is included, a little higher when multiplying is omitted. This corresponds to the result that correlations by the best records method are less reliable in multiplying than by the median method. Correlations with word-building are similarly occasionally lower and are (see table 3) less reliable. Correlations during the eighth practice period are a little higher for the "best" series. For three tests the reliability is greater. "Best" records show, then, neither a higher nor a lower correlation than median records. The same progression is observable—an increase up to the fifth or sixth periods, followed by a slight decrease.

Still another measure may be used. We may do away with the eight divisions of the practice period and consider each subject's highest point only. The best score of each individual in each test, whether this record is attained in the beginning, middle, or at the end of the practice series has been correlated with her best record in every other test. The variability of such a record cannot be determined. Presumably it is large, since such a score would be subject to many possible errors. Accidental factors would lower the correlation below its true amount. The table follows:

TABLE NO. 16

Correlation of each Test with every other at Highest Point reached by each subject.

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.		52	52	05	27
Tap.	52		33	-19	23
Add.	52	33		36	32
Mul.	05	-19	36		01
W. B.	27	23	32	01	
Av.	34	22	38	06	21
Av. (without M.)	44	36	39		27

The average of correlations with multiplying is .24, without multiplying it is .37. The .24 is equivalent to the average for the correlation of medians of the third and fourth division, the .37 is nearest the .39 obtained by averaging correlations in the fifth and sixth divisions, when multiplying is omitted. Uncorrected correlations of best records show, with the exception of multiplying, a relatively high correlation,

about equal to that obtained by the median record method at its highest point.

The cause of the increase and decrease of correlations, observed both in median and best records, of the equal reliability of median and best records, of the relatively high correlation of very best records when compared with their probably unreliability, is difficult to find. Rise and fall in coefficients have been attributed to improvement through practice, which increases correlations, up to the time of the entrance of another factor, fatigue. Jones believed that a decline in correlation was due to the elimination of the adaptation factor which raised the initial correlation above its true amount. Hollingworth's results showed a steady increase in coefficients which he attributed partly to lessened variability, partly to change in the nature of the tests, and partly to the fact that final tests were more representative of ultimate capacity than were initial trials.

In the present results a parallelism between the change in size of intercorrelations, the change in size of adjacent correlations, and the amount of gain made by practice is at once apparent. Up to the fifth or sixth trials individual and average scores (see Chapter 3) show a great improvement. After this there is an improvement level, or in some cases retrogression. Adjacent correlations are lowest at the beginning of practice, higher at the very end, but greatest of all at the fifth or sixth practice period. Similarly intercorrelations show an increase and subsequent slight decline.

The coincidence of the changes would tend to suggest one progression as the cause of the others. Obviously the increase in individual scores is the primary factor. Without change in records there can be no variation in correlations based on these records. While great progress is being made, variability is great and intercorrelations low, at the time of maximum achievement variability is least and intercorrelation highest, where achievement is somewhat reduced there appears a somewhat greater variability and a somewhat lower intercorrelation.

A change in variability may be suggested as the entire cause of the change in intercorrelation. If this is the case, then when correlations are corrected for attenuation, the differences should disappear and beginning, end, and middle correlations should be approximately equal. Table 14 gives

the corrected coefficients for the second, fifth and eighth periods.

TABLE NO. 17

Corrected Coefficients (Median) 2nd Period.

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.	x	40	36	–16	25
Tap.	40	x	–27	–03	27
Add.	36	–27	x	69	51
Mul.	–16	–03	69	x	44
W. B.	25	27	51	44	x
Av.	24.6				
Av. without M.	28				

5th Period

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.	x	69	37	–05	39
Tap.	69	x	27	12	48
Add.	37	27	x	93	55
Mul.	–05	12	93	x	10
W. B.	39	48	55	10	x
Av.	38.5				
Av. without M.	46.0				

8th Period

C. N.	x	51	40	09	18
Tap.	51	x	27	11	–08
Add.	40	27	x	78	47
Mul.	09	11	78	x	21
W. B.	18	08	47	21	x
Av.		29.4			
Av. without M.	29.2				

Coefficients still vary. The average with multiplying progresses from 25 to 39 to 30, without multiplying from 28 to 46 to 29. This is to be compared with the series 19, 33 and 25, and 21, 41 and 25 of the uncorrected medians.

	Δ2 & 5	Δ5 & 8
Median Uncorrected (with M)	14	–8
Median Corrected (with M)	14	–9
Median Uncorrected (without M)	20	–16
Median Corrected (without M)	18	–17

The difference then, does not disappear. Coefficients are on the average, raised but the gross difference between the averages of the correlations of the different periods remains unchanged. The increase in inter-correlation cannot, then, be attributed wholly to a change in variability.

Other suggestions pointing to the activity of some other factor, would be the equal reliability of best as compared with median records, and the somewhat superior inter-correlations obtained by working with very best scores.

We must suppose that approach to an improvement level gives not only decreased variability but equalizes the effect of previous practice and enables the individual's ultimate capacity to appear more plainly. We might say that best scores in each period are no more variable than determinations based upon more than one score, just because the former are more representative of final ability. Similarly very best records which seem to be the best evidence of what ultimate achievement might be, show, even though uncorrected, high inter-correlation. Not accuracy alone but the quality of representing ultimate capacity, seems to be the characteristic of records which produce higher inter-correlations.

In this connection one should note the meaning of the coefficient of reliability which has been used. It is a measure of the change in relative position from one median record to the next. Where change is greatest, as at the beginning of practice this coefficient will be lower, not only because records vary more but because some individuals are gaining more, some less. Where there is relative stability the coefficient is high, where some retrogression it is lower again. The use of it for correction would seem to raise initial correlations above their true amount, for not only variability but different amounts of improvement are excluded. At the end of practice a high adjacent coefficient is caused not only by less variability but by greater equality of gain from practice.

It may be, then, that correlations tend to increase as final ability is approximated. Final capacity probably best represents true capacity, ability freed from such influences as accidental initial high or low performance, variability of a few trials, or inequality of previous practice. True measures of individuals would presumably show a positive and fairly high correlation. A number of the negative correlations found may be due to the inaccuracy of the measures. But again, it is important to remember that these correlations are not, except in a few cases, remarkably high. Occasionally the change is less than the probable error of the median coefficient. That chance is not entirely responsible for the change is evidenced by the constancy of direction of the variation. That the correlations are as high as they are, in view of the relative homogeniety of the subjects is a fact pointing toward a possible finding of great positive relationship between abilities tested in a random sampling of humanity.

An examination of individual coefficients should be added to determine if possible, the functions operative in the increase and to suggest as well as so few tests can, the different degrees of relation between traits. The median uncorrected table will be used throughout.

The correlation between color-naming and tapping is always positive and high. It increases up to 67 in the sixth division. Color-naming and adding begin with a correlation of 20 which increases up to 41. Color-naming and multiplying begin with a negative correlation which remains less than twenty throughout the series. The correlation between color-naming and word-building begins at 21, goes down to 08, but reaches finally 41. Tapping and adding begin at 24, correlate twice negatively, finally reach 36. Tapping and multiplying begin at –16, remain minus till the fifth trial where they become slightly positive, never becoming, however, greater than 10. Tapping and word-building begin at 12, become negative, finally reach 40 after which they are again negative. Adding and multiplying always have a high positive correlation which begins at 30 and increases to 78. Addng and word-building change from 24 to 46. Word-building and multiplying increase from 20 to 32 but remain irregular throughout.

Results from so few tests are of little value for the solution of general theoretical problems concerning the relation between traits. A few facts may be observed.

The importance of identical elements and of content is notable in the high correlation of adding and multiplying. In spite of the unrelability of the multiplying test, uncorrected coefficients of adding and multiplying are high—due probably both to native ability and special training in a definite content, and to the identity of the elemental operations. A low correlation of multiplying with color-naming and tapping is observed. Is multiplying possibly an ability of a higher level, and is the lack of correlation due to difference in the kind of performance? Multiplying and word-building, both of which are tests of the analytic and selective variety show a somewhat higher correlation. Perhaps the correlation would have been higher still had multiplying been aways a test of analysis, had the determination been more reliable, and the amount of practice greater. We have a very high correlation between color-naming and tapping—functions of the associative and sensori-motor level.

The averages with multiplying show that adding has the highest correlation with all other tests, color-naming the next, then word-building, multiplying, and tapping. When correlations with multiplying are omitted the order of greatest correlation is color-naming, adding, tapping, and word-building. We might attribute the high correlation of color-naming and adding to the relative lack of variabilty of the tests and to the fact that in these an improvement level was approximated.

We can make no conclusion as to general theory of the cause of the observed relationship. Correlations do seem to depend upon identity of elements, they are higher in the case of multiplying for a test of the same, than for a test of other levels. They are influenced by the amount of variability of the test, and, apparently, by approximation to proficiency.

SUMMARY

The uncorrected coefficients obtained in the present investigation confirm the suggestion offered by Hollingworth that practice increases the amount of inter-correlation.

The increase in size of coefficients coincides with increased individual and average scores which are the result of practice, and with decreased variability.

When coefficients are corrected for attenuation the difference does not disappear.

This increase may then, be attributed, to the fact that true measures of ultimate capacity are being approximated.

This is borne out by the fact that a measure which best represents ultimate capacity shows, in spite of its being probably subject to incidental variation, a relatively high correlation.

Those tests in which an improvement level is approximated and which are simpler and more reliable have on the whole, higher inter-correlations.

On the whole, tests of the same level correlate a little closer than do tests of different levels.

CHAPTER FOUR

THE RELATION OF IMPROVEMENT TO INITIAL ABILITY

If it were true that correlations between tests increase with approach to an improvement level, then we might say that practice with improvement tends in one sense, at least, to increase or to emphasize indvidual differences. Instead of finding one person capable in one direction, one in another, we might discover that some people were good in most lines, others poor in most lines, others mediocre. Taking the whole range of abilities into consideration it might be shown that the average status of each individual, when the compensating influence of special training was equalized, was more unlike the average status of each other individual after, than before practice gain.

We might, on the other hand, observe one instead of a number of functions, noting the influence of practice in that one ability upon differences as displayed in that trait. One way of approaching this problem would be to inquire—who improves most, the initially poor or the initially good individual? If the one who is most efficient in the beginning, makes the greatest gain, then practice would seem to increase differences. If the one who is originally poor, improves the most, then persons would seem to approach each other, as repetitions are increased. A finding of the first sort would seem to show that native ability evidenced, perhaps, by the first record made, rather than environment, in the form of special training, was the factor responsible for proficiency. A result of the second sort would seem to mean that, since the equalization of practice would decrease differences, inequality of training was the cause making for variation.

A number of studies have been made which deal more or less directly with this problem. Among those which bear definitely on the question, the investigations of Thorndike, Whitley, Wells, Brown, Kirby, Donovan, Hahn, Chapman and Dallenbach are of greatest interest.

Thorndike,* in 1908, investigated the effect of practice in the mental multiplication of one three-place by another three-

*Thorndike—The Effect of Practice in the Case of a Purely Intellectual Function, A. J. P. Vol. 19.

place number. Twenty-eight individuals solved ninety-six examples each. The subjects kept their own time records and errors were afterwards transmuted into time. Equality of practice, therefore, means an equal amount of work performed. In this function the physiological limit is far below the ability reached by any individual although, as Thorndike states, two persons seem to reach a plateau. The table showing the improvement made by individuals differing in initial ability is repeated here.

TABLE NO. 18

The Effect of Equal Amounts of Practice upon Individual Differences in the Mental Multiplication of a Three-place by a Three-place number.

	Hours of Practice	Amount done per Unit of Time. 1st 5 Examples	Last 5 or 10 Examples	Gain	Percentage of Correct Figures in Answer. 1st 5 Examples	Last 5 or 10 Examples	Gain
Initial highest 5 Individuals	5.1	85	147	61	70	78	18
Initial next 5 Individuals	5.1	56	107	51	68	78	10
Initial next 6 Individuals	5.3	46	68	22	74	82	8
Initial next 6 Individuals	5.4	38	46	8	58	70	12
Initial next 5 Individuals	5.2	31	57	26	47	67	20
Initial next 1 Individual	5.2	19	32	13	100	82	—18

The table shows that the best at the start made a larger gain in amount done per unit of time and an approximately equal gain in accuracy, with the exception of one individual.

A study of practice in adding seemed at first to bring opposite results.* The correlation between initial ability and percentile improvement was a negative one, $-\frac{1}{4}$, and for gross improvement (by my computation) –68. The ratio of highest to lowest individuals as given by Thorndike in the original article changes as shown in this table.

TABLE NO. 19

	Ratio of Highest to Lowest Individual				
	1st 19th	2nd 18th	3rd 17th	4th 16th	5th 15th
In first practice period	2.49	2.25	1.92	1.82	1.46
In second " "	3.36	1.99	1.87	1.62	1.46

*Thorndike.—Practice in the Case of Addition A. J. P. Vol. 2, pp. 483-486.

Thorndike says, "The differences amongst individuals in the ability to add seem to be due in larger measure to differences in environmental influences: For equal practice does reduce a little the relative or percentile differences within our group." And farther on "There is, of course, no essential conflict between this result for addition and the opposite result for mental multiplication with two three place numbers. The same theoretical view which would expect mental span and ability to manage very complex relationship in a given field to be increased by practice in close dependence upon original capacity would expect particular associative habits such as thinking of thirteen upon seeing 4, 7 and 2 in a column to be increased by practice in less close dependence upon original capacity."

These results were obtained and computations made in terms of amount of time required to perform a definite task. In 1914 the results are again reported.* There the average number of additions per 5 minutes, corrected for errors, is the score used and each of the nineteen subjects is given an approximately equal time for practice. The table follows.

TABLE NO. 20

The Effect of Equal Amounts of Practice upon Individual Differences in Column-Additions of One-Place Numbers.

	Average Number of Additions per 5 minutes corrected for errors			Average time spent in practice from mid-point of first test to mid-point of last test (in minutes)
	First Test	Last Test	Gain	
Initially highest 6 Individuals	297	437	140	40
Initially next highest 6 Individuals	234	345	111	49
Initially lowest 7 Individuals	167—	220+	54	46

As the facts are here presented, it is apparent that the initially best gain the most.

When, then, decreased variability after practice and lack of correlation between initial ability and improvability computed in terms of time required are used as measures of the effect of practice, individual differences seem to decrease with repetition. When the subjects are divided into groups on the basis of their initial ability, when the amount measure is used and the time of practice made approximately equal, the average improvement made by each of these groups seems to show that equal amounts of practice are effective in increasing individual differences.

*Thorndike—Educational Psychology, Vol. 3.

The results of Whitley's[1] correlations are in substantial agreement with those of Thorndike. After an analysis of methods of measuring changes in which she shows the different results which may be obtained by measuring gain in time, in amount, in per cent, and by two variations, (the Smythe Johnson and Amberg) of the percentile method, she gives results obtained from five tests practiced by nine women for twenty days. The tests were the familiar curved maze, mental multiplication of two three place numbers, a sorting test, a cancellation test and a lifted weight test. Correlations between position at start and at the end of practice, and position at the start and average position, or gross gain or percentle gain, were:

TABLE NO. 21

	Pos. at start & at Finish	Pos. at start & Av. Pos.	Pos. at start & Gross Gain	Pos. at start & Per. Gain
Lifted Weights	.27	.45	-.98	-.65
Cancellation Tests	.72	.58	-.90	-.38
Sorting Test	.56	.58	-.92	-.36
Mul. Test	.44	.58	-.63	-.52
Maze Test	-.21	.33	-.95	-.90

There is a low positive correlation between initial and final ability in all these tests but the maze and a slightly higher correlation between position at the start and average position. In all tests there is a relatively high negative correlation between gain and rank in the first trial. Whitley says of these results: "Individuals with low standing can and do improve the most judging objectively though even so, they may not in conveniently measurable periods of time overtake those whose standing was high at the beginning."

In concluding her discussion of the improvement curve she says: "All functions do approximate one typical law for changes in improvement though individuals may tend to differ in variability. From this point of view practice must tend to make people more alike."

Thorndike in 1914 says of Whitley's results:[2] "The same effect (meaning an increase of difference due to practice) appears though less emphatically in the case of Whitley's nine individuals in a similar experiment. The four who were most efficient at the start made a greater average gain from equal

[1]Whitley—An Empirical Study of Certain Tests for Individual Differences, Archives of Psychology.

[2]Thorndike.—Education Psychology.—Volume 3.

practice than the four who were least efficient." Again averaging the records of groups of persons gives results apparently contradictory to those obtained by the correlation method.

Wells,[1] (1912), made a study of the effect of repetition in an adding and a cancellation test, using five men and five women as subjects. His results, based on a comparison of the improvement curves obtained, show that when absolute gain is considered, both men and women shows an association of high initial ability with possibility of great improvement. It is notable, however, that all of these comparisons are subject to exceptions—an individual whose record is good at first may make little gain, or an individual whose original ability is poor may make a surprisingly large gain from practice. Measures by relative gains show "no negative relation" with gross performance. The Table which Wells gives showing the orders obtained by the different measurements is given here.

TABLE NO. 22

Well's Table (Individuals designated by Names of Colors)

Initial Performance		Absolute Gain		Relative Gain	
Addition Test (Men)					
Blue	1	Blue	1	Red	4
Brown	2	Brown	2	Green	5
Orange	3	Red	4	Brown	2
Red	4	Green	5	Blue	1
Green	5	Orange	3	Orange	3
Additional Test (Women)					
Blue	1	Red	3	Red	3
Green	2	Blue	1	Blue	1
Red	3	Green	2	Brown	5
Orange	4	Orange	4	Orange	4
Brown	5	Brown	5	Green	2
Number-Checking Test (Men)					
Green	1	Brown	2	Brown	2
Brown	2	Green	1	Red	4
Blue	3	Red	4	Green	1
Red	4	Blue	3	Orange	5
Orange	5	Orange	5	Blue	3
Number-Checking Test (Women)					
Red	1	Blue	4	Brown	5
Green	2	Orange	3	Orange	3
Orange	3	Green	2	Blue	4
Blue	4	Brown	5	Green	2
Brown	5	Red	1	Red	1

[1]Wells.—The Relation of Practice to Individual Differences, A. J. P. Vol. 23, pp. 75-88.

The figures whch I have added giving the initial rank of the individuals at the right of the columns shows the changes even more clearly. Red (Number-Checking-women) changes from first to fifth position when initial ability and gross per cent gain are compared. Orange and Brown (Number-Checking Men & Addition Women), on the other hand, are fifth in two cases. The table shows both negative and positive results.

Wells adds to his comparison of improvement curves, a discussion of the value of different measures of improvability, favoring absolute measures. Factors in improvement seem to be, in the order of their influence in the present tests, (1) a difference in the individuals (resp. functions) plasticity, i. e., ability to profit by practice, (2) a difference in the actual amount of practice experienced and (3) constitutional factors, independent of plasticity, in the nervous system. He concludes:

"The more significant results would seem to be the indication that a superior performance at the beginning of special practice is not necessarily nor even probably attained at the sacrifice of prospects for future improvement. A high initial efficiency may carry with it as much as or more prospect of improvement under special practice as a low one. It was not because the favored individual had had more of the general experience enabling him to meet the experimental situation better but because he possssed the native abiity to profit more by such experience, general and special, past and future. Not practice, but practicility, is responsible for the superior position of such an individual and, in broader aspect, not education but educability."

Warner Brown* who studied improvement in card sorting, had twenty-six individuals make four or eight trials a day for thirteen days. In addition to results obtained concerning interference of habits he reports data on the relation between initial and final speed and improvability. His records showed (1) a high positive correlation between speed on the first day and speed on any succeeding day (2) a negative correlation between speed on the first day and ability to profit by practice, (3) a negative correlation between speed at the end of thirteen days and improvement made during this time." There are a

*Brown.—Interference in Card Sorting. University of California Publications.

few exceptions to the second finding—a few persons were both quick at the beginning and able to gain a great deal from practice. "Apart from these exceptional cases," Brown says, "it seems on the whole unquestionable that those persons who begin slowly will generally manifest a greater capacity for improvement (relative as well as absolute) than those persons who begin work at a high rate of speed. The effect of practice in this complex process is to flatten out the differences between individuals." On the other hand, he says: "Large practice gains do not insure the attainment of superior skill, on the contrary they serve as a rule only to reduce the handicap of those persons who are slow at the beginning."

Hahn and Thorndike[1] give the following table for amount of improvement made by school children in arithmetic. The numbers in the table refer to number of examples correctly solved. The numbers in parentheses designate the number of individuals considered.

TABLE NO. 23

Amount of Improvement in Relation to Initial Ability.

Initial Ability	Grade 4	5	6	7	Average of Grades
0-6	12.3 (10)	10.7 (7)	9.0 (29)		10.7
7-12	7.6 (9)	9.7 (12)	10.1 (9)	15.0 (2)	10.6
13-18	13.4 (5)	7.6 (8)	12.3 (6)	17.5 (2)	12.7
19-24	6.7 (3)	19.5 (4)	11.0 (6)	20.4 (9)	12.3
25-30		12.3 (4)	1.5 (2)	23.4 (6)	12.4
31-36		54.0 (1)	8.5 (2)	16.4 (5)	
37-42				9. (1)	

The authors conclude from this table that individual difference increase with practice since the initially more proficient gain on the average, more.

The variability of the averages obtained, is, however, very great. "Initial ability 25-30," varies from 1.5 to 23.4, "19-24" from 6.7 to 20.4, "31-36" from 8.5 to 54.0.

Data obtained by Donovan and Thorndike (1913)[2] gave similar results. The average gross gain in number of examples done for the four initially least proficient of 29 fourth grade pupils was 5.3, for the next four, 10, the next four, 5.3, the next seven, 6.6, next four, 8.3, next five, 6.8, next two, 11.

The authors say: "These results, showing so little power of equal addition to training to reduce individual differences, make it improbable that a very large function of the difference

[1]Hahn and Thorndike.—Some Results of Practice in Addition Under School Conditions. Journal of Education Psychology, Vol. 5 No. 2.

[2]Donovan & Thorndike.—Improvement in a Practice Experiment Under School Conditions, A. J. P. Vol. 24, pp. 426-428.

found among school children can be greatly attributed to difference in amount of training." One does not find, in this table, however, a definite tendency for the best persons at the start to improve more. The most that we can safely say is that there is an equal improvement of best and poorest.

Kirby's curves,[1] drawn on the basis of the performance of about 700 school chidren practising addition and division for equal amounts of time show a positive relation between gain and initial ability.

Chapman's[2] results for 22 individuals, five tests, show a negative correlation for the simpler functions which are nearer their physiological limit, a positive relation for the more complex functions.

TABLE NO. 24

Correlation between Initial Scores and Improvement.

	Initial Score	(1 2 3)
Column I	Improvement	(6 7 8) – (1 2 3)
Column II	Improvement	(9 1 0) – (1 2 3)

	Col. I Correlation	Col II Correlation
Color	03	–.14
Cancellation 2	06	05
Opposites	–.09	–.19
Addition	.44	.34
Multiplication	28	.41

The correlations between initial and final position are—color 87, cancellation (2) 75, cancellation (3) 85, opposites 50, addition 96, multiplication 87—all it should be observed, are high and positive. Chapman concludes: "It appears that in complex functions an individual who has gained high efficiency by previous general training will also improve correspondingly during the practice period, whereas in simpler functons as usually scored, initial efficiency probably bears little relation to improvability. If the method of scoring were altered so as to weigh improvement which is made as the subject approaches the physiological limit, there is little doubt that high correlations would be obtained in the simpler functions."

One recent study of normal and backward school children gave important results. There the function studied (Dallen-

[1]Kirby.—Practice in the case of School Children, Teachers College, Columbia University Contributions to Education No. 58.

[2]Chapman.—Individual Differences in Ability and Improvement, and their Correlations.—Col. U. Contributions to Education.

bach 1919)* is visual apprehension. Both normal and backward children were divided into three groups on the basis of their performance in initial tests. The amount of improvement computed for each group is shown in the following table.

TABLE NO. 25

Normal Group		Feeble-Minded Group	
Initial Ability	Improvement	Initial Ability	Improvement
42.72	13.74	48.5	6.12
35.32	13.43	41.26	3.14
28.56	15.71	34.31	2.31

The feeble-minded group were initially more proficient. They improved less, however, so that the normal group finally surpassed them. Dallenbach says: "The feeble-minded and poorer third of normal group continue to improve whereas the superior and mediocre subjects of the normal group show a large initial increase and then a long plateau." It seems improbable that a function like this should be near its physiological limit, yet when abnormal are compared with normal, or initially poorer normal with initially better normal—the individuals best at the start are seen to improve least. When comparisons are made within the feeble-minded group the opposite effect is observed.

In this problem then, results and especially conclusions drawn from them differ widely. The apparent contradictions should be due, it seems, to the fact that the expermientors studied different functions, permitted different amounts of repetition, used different methods of measuring gain and equating practice. Occasionally different conclusions may be drawn from the same experiment. A correlation figure may be urged in support of one assumption while an observation of the trend of average results may lead to another conclusion. We find, however, a certain amount of agreement. With the exception of Whitley's results, experiments on mental multiplication, adding, dividing, show in general, a positive, or, at the most, a zero relation between initial ability and gain. Improvement in tests like color-naming, opposites, cancellation, card-sorting, visual apprehension seems normally to be negatively related to initial proficiency. Those performances in which practice is measured in terms of equal time spent (as in Kirby's, Donovan and Thorndike's, Hahn and Thorndike's experiments) show more frequently a positive correlation than

*Dallenbach—A Test for Visual Apprehension Jour. Ed. Psy., 1919.

do those in which equality of practice means the reaching of a certain result a definite number of times (as in all color-naming tests, Whitley's and Brown's investigations and Chapman's opposites.) Exceptions to this are Dallenbach's visual apprehension and Chapman's cancellation test. In both cases, though equal practice meant equal time spent, the correlation was negative. Conclusions seem to depend upon the nature of the function studied and the measure of equality of practice used. The problem of the measure of improvement is similarly of importance.

As Thorndike, Whitley, Wells and others have pointed out, the question of the measurement of gain is a complex one. Theoretical difficulties are involved, as the problem as to whether gains made at different parts of the practice curve may be called equivalent. Even if we confine ourselves to actual time, amount, or percentile figures, asking not about relations of gain considered in the light of their frequency, difficulty, etc., but about gains compared on the basis of some numerical scores, questions of the interval over which improvement shall be measured and of the score which shall be used, remain. It might be well instead of choosing one measure of gain and basing all conclusions on that, to proceed empirically, to try out a number of methods and to compare the results obtained. Whitley has demonstrated the different conclusions reached when theoretical scores are treated by different methods of computing. A like study, including time, amount, and per cent methods, and using two different intervals instead of one, might be made upon actual scores.

In the present investigation we have measured initial ability by the record made in the first trial or by the median record of the first three or the first five trials. Each method has certain advantages and certain difficulties. The second is more accurate than the first in that the record is obtained from more than one trial. This gain in accuracy may be balanced by a loss in definiteness due to the introduction of another factor—improvement in the tests. When we take into account not only the first but the second and third or perhaps the fourth and fifth trials—we are measuring not only the first attack at the problem—immediate adaptability—but the progress made during three or five trials. By employing both methods we make possible a comparison of the two and a correction of one by the other.

Final ability has also been measured in two ways by using the median of the last three trials or the best record made. The former measure is more accurate than a ranking by the last trial alone. The effect of especially high or low records at the end—due perhaps to the excitement of finishing the series or the variability occurring there are partly eliminated by this method. Learning, which would make the score of the first trial differ greatly from the median of the first three or five trials, would be of no practical importance at the end of the period of work when change is negligible. The interval between the median of the first and of the last three trials was the first basis upon which improvement was scored.

As a corresponding point to initial ability as measured by first trial, final ability as measured by the very best record wherever it occurs was chosen. The main objections to the choosing of such a point are—(1) the large part that chance may play when measurement is based on only one trial, (2) the theoretical objection that since this point is reached by different individuals at different parts of the practice curve that the amount of practice is in no sense equal—and that discussion of different amounts of improvement upon the same amount of practice is irrelevent. To the objection of possible inaccuracy we have no adequate answer. Factors such as fatigue, lack of interest, distraction during the test are mentioned by writers on practice as possible causes of inaccuracy. These would tend to make individual ability seem less than its real amount. The only chance errors which would operate to raise records beyond the ability of the individual would be errors due to apparatus or scoring methods, or "accidents in the nervous system" which might make a subject show a great proficiency, never again equaled. That the second sort of errors are less frequent than the first seems in the absence of evidence to the contrary, at least, possible. Best records might even approximate a representation of final ability when the errors of the experiment were eliminated. That such scores are fairly reliable seems to be shown by their relatively high inter-correlation (see Chapter 4.) At any rate, if they are less reliable (and if all other conditions are equal) this method will reduce all coefficients toward zero, making the positive correlations less positive and the negative less negative.

The second objection may be met by adding another record to the scores at the beginning and at the end, and by making

a change in the nature of the problem that we are discussing. The additional record is one that I have called "when reached." It is a measure of the time at which the best record is attained by each individual. The theoretical way of meeting this difficulty consists in changing our problem for the time from a discussion of improvement upon equal amounts of repetiton to a discussion of improvement based on equal opportunity for learning. Both kinds of equality, it should be recalled are included in the experiment. In the tapping and word-building tests there is an equal amount of time spent, in the other tests an equal amount of products produced. The practical aspects of the problem seem to remain almost unchanged. We give groups of subjects an equal amount of practice. The problem is not—as in the former case—who at the end of practice has attained the highest level of efficiency? but rather—who at any time in the course of the study has shown herself able to reach, by the aid of her original nature and the gain from as much practice as she may have had—a high score? A person who reaches her best score early in the practice period and whose improvement is measured on this basis, has had a very small amount of practice in comparison with her fellows at the time the measurement is made. She has, in the end, however, equal opportunity for making progress. She is not penalized by measuring her improvement on the basis of a few trials. Her record is on the contrary better than it would otherwise have been. Her lack of ability to maintain her high standing is not taken into account. Whether or not this lack of ability is a significant thing or merely an accident due to the inaccuracy of the experiment—is a point to be considered after the different measures of gain have been compared. Our two problems, based on different intervals of measurement are: "Which individuals, with equal amounts of practice, will be, at the end, at the highest level efficiency?" and "Which individuals, being given equal opportunity to dispay their ability, will best demonstrate the possibility of eventual high achievements?"

The two intervals—the difference between the median of the first three and last three trials—or the difference between the first and the best trials were used then as our basis for calculating improvement. Four methods of measuring gain were employed for both the first and the second periods. The first interval is measured by—(1) the difference between the gross

time required, (2) the difference between the gross amount of work performed in a given time, (3) the per cent which the gain measured by either of these methods was of the initial abiity, (4) or by the method of changes in ranks. An example will make these scores clear.

Let us suppose that "A" in the adding test has an initial score of 103 seconds required to add the first four columns and a final score of 79 seconds for time required to add an equally difficult four columns. The gross difference in time (first method) is 24. The gross difference in amount is obtained by transmuting the scores 103 and 79 into terms of amount done per unit of time. The difference between the reciprocals of the two numbers gives us .003. The ratio, gain divided by first trial, is different by the time and by the amount method —in one case .233 in the other .309. This difference by our scheme of correlating does not concern us—for though the actual percentage amounts may be different, their relative order by the time and amount method remain unchanged. In using the method of ranks we need to arrange all the subjects in an order of merit for initial and final ability and to use as a measure of gain the difference between the first and last positions of each individual. The person whose gain was the largest has made the largest step up on the scale—the person whose gain was the least has gone down the largest amount on the scale. All steps it should be noted, are assumed to be equal. A gain from second to first place is considered equivalent to a change from twenty-third to twenty-second positions. The value of this assumption can be tested later. In our illustrative case A—who was 20th at first became 16th in the final ranking. Her gain then was 4.

Using the second interval—difference between first and best trials—as a basis of measurement we obtain four figures —difference in time required, difference in per cent of gain, difference in relative position and difference in time at which the best record was made. The first three measures are similar to those already employed. The measure "gain in amount" was omitted, because the labor of turning time into amount seemed to over-balance the possible advantage of another measure of improvement. The possible trend of correlations by this method may be inferred perhaps by applying the results obtained when the difference between the median of the first and second trial was used, to the actual time records se-

cured from the second interval. The measure, "time at which best record is made," was obtained in the following way: The subject who made her best record earliest in the practice was considered to have made the most significant gain. In our illustration, "A" achieved her best record at the twentieth trial—her ranking is higher by this method than "B" whose best record came at the twenty-first trial.

The following tabular arrangement will give a better understanding of the measures used.

1. Basis of measurement—progress made between median of first three and last three trials.

A. Improvement measured by difference in gross time required.

B. Improvement measured by difference in gross amount done per unit of time.

C. Improvement measured by ratio, gain measured by gross methods divided by median of first three trials.

D. Improvement measured by gain in relative position.

II. Basis of measurement—progress made between initial trial and best trial.

A. Improvement measured by difference in gross time required.

B. Improvement measured by ratio, gain measured by gross method divided by first trial.

C. Improvement measured by gain in relative position.

D. Score used—time of reaching best record.

The two problems—measurement of gain and relation of improvement to initial ability may best be attacked (1) by considering the general trend of relationship, the result obtained when all the different measures of gain were correlated with the measures of initial ability, (2) by examining the agreement between the measures of gain—both by (a) comparing the correlations with initial ability obtained by using different measures of gain and (b) by comparing the tests themselves— using as a basis for this comparison, the observed relationship between original ability and improvability. Table 26 presents the data on which the conclusions are based. It shows the correlation of scores signifying improvement, with each other and with initial proficiency. In this table are included not only the results of the present investigation but a computation based upon a study made by Professor Holling-

worth in 1914.* In this experiment thirteen individuals practised seven tests, repeating each test two hundred and five times—until a definite improvement level was reached. All

TABLE NO. 26

Correlation between Initial Score and Various Measures of Gain. Investigation A—Gain Measured from Medians.

Color-naming.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	–31	–02	–24	–39
Time	–31	x	93	98	87
Amount	–02	93	x	97	96
%	–24	98	97	x	96
Rank	–39	87	96	96	x

Adding.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	–11	50	16	–12
Time	–11	x	78	93	91
Amount	50	78	x	91	74
%	16	93	91	x	88
Rank	–12	91	74	88	x

Multiplying

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	–84	30	–17	50
Time	–84	x	13	60	71
Amount	30	13	x	85	88
%	–17	60	85	x	66
Rank	50	71	88	66	x

Tapping.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	–15	–09	–15	–38
Time	–15	x	99	99	89
Amount	–09	99	x	99	87
%	–15	99	99	x	88
Rank	–38	89	87	88	x

Word-Building.

	Median 1st—3	Time	Amount	%	Rank
Median 1st—3	x	–26	–05	–11	–24
Time	–26	x	88	97	92
Amount	–05	88	x	96	92
%	–11	97	96	x	90
Rank	–24	92	92	90	x

Investigation A—Gain Measured from 1st Trial to Best Trial.

Color-Naming.

	1st Trial	Time	%	Rank	When
1st Trial	x	–57	–40	–39	–11
Time	–57	x	94	80	–12
%	–40	94	x	87	–20
Rank	–39	80	87	x	–36
When	–11	–12	–20	–36	x

*For further description of the tests see Hollingworth's "Influence of Caffein upon Efficiency." Archives of Psychology, No. 22.

Tapping.

	1st Trial	Time	%	Rank	When
1st Trial	x	06	−18	−44	50
Amount	06	x	99	85	−21
%	−18	99	x	89	−27
Rank	−44	85	89	x	−41
When	50	−21	−27	−41	x

Adding.

	1st Trial	Time	%	Rank	When
1st Trial	x	−85	−68	−52	−07
Amount	−85	x	96	83	02
%	−68	96	x	91	−01
Rank	−52	83	91	x	−42
When	−07	02	−01	−42	x

Multiplying

	1st Trial	Time	%	Rank	When
Ist Trial	x	−94	−66	−48	−03
Amount	−94	x	78	61	20
%	−66	78	x	90	15
Rank	−48	61	90	x	18
When	−03	20	15	18	x

Investigation B—Gain Measured by Difference Between Median of 1st and Last 5.

Color-Naming.

	Median 1st—5	Time	Amount	%	Rank
Median 1st—5	x	−73	−29	−57	−36
Time	−73	x	81	97	83
Amount	−29	81	x	87	83
%	−57	97	87	x	90
Rank	−36	83	83	90	x

Opposites.

	Median 1st—5	Time	Amount	%	Rank
Median 1st—5	x	−98	−35	−83	−63
Time	−98	x	45	92	71
Amount	−35	45	x	66	90
%	−83	92	66	x	76
Rank	−63	71	90	76	x

Calculations

	Median 1st—5	Time	Amount	%	Rank
Median 1st—5	x	−92	−50	−69	−65
Time	−92	x	72	88	83
Amount	−50	72	x	95	94
%	−69	88	95	x	98
Rank	−65	83	94	98	x

Tapping

	Median 1st	Time	Amount	%	Rank
Median 1st	x	−03	10	−03	−50
Time	−03	x	98	99	82
Amount	10	98	x	98	86
%	−03	99	98	x	81
Rank	−50	82	86	81	x

Discrimination

	Median 1st	Time	Amount	%	Rank
Median 1st	x	−53	−53	−53	−36
Time	−53	x	1.00	1.00	78
Amount	−52	1.00	x	1.00	78
%	−53	1.00	1.00	x	78
Rank	−36	78	.78	.78	x

Crossing Test (Cancellation)

	Median 1st	Time	Amount	%	Rank
Median 1st	x	–02	27	20	–43
Time	–02	x	90	64	85
Amount	27	90	x	99	75
%	20	64	99	x	74
Rank	–43	85	75	74	x

Three Hole Test

	Median 1st	Time	Amount	%	Rank
Median 1st	x	–89	–08	–66	–32
Time	–89	x	51	89	58
Amount	–08	51	x	78	87
%	–66	89	78	x	82
Rank	–32	58	87	82	x

measurements were made in terms of time required to perform a given task. The color-naming, tapping, and calculation tests of this study are comparable to the color-naming, tapping and adding tests of the present investigation. The records here have been treated by the first method, the interval being that between the median of the first and of the last five trials. This study will be referred to as Investigation B. Table 27 is an extract from the first table, presenting some of the same facts in a different order. Correlations between measures of gain have been omitted and the relation of improvement to initial ability only is shown. An inspection of this table will give an idea of the general trend of relationship.

TABLE NO. 27

Correlation Improvement and First Records.

I. Investigation		A—Measured by Median.			
Test	Time	Amount	%	Rank	When
C. N.	–31	–02	–24	–39	
Tap.	–15	–09	–15	–38	
Add.	–11	50	16	–12	
Mul.	–84	30	–17	50	
W. B.	–26	–05	–11	–24	
II. Investigation		A—Measured by interval of 1st and best trials.			
C. N.	–57		–40	–39	–11
Tap.		06	–18	–44	50
Add.	–85		–68	–52	–07
Mul.	–94		–66	–48	–03
III. Investigation		B—Measured by Medians.			
C. N.	–73	–29	–57	–36	
Opp.	–98	–35	–83	–63	
Calcu.	–92	–50	–69	–65	
Tap.	–03	10	–03	–50	
Discr.	–53	–53	–53	–36	
Cross-Test	–02	27	20	–43	
Three Hole	–89	–08	–66	–32	

A definite tendency of the correlations is at once apparent. The correlation of gain, measured in any of the five ways, with initial ability is negative in fifty-six cases out of the sixty-four. It is above minus 30 in one-half of the cases, above minus 10 in all but seven of the negative cases.

Exceptions to this tendency toward negative correlation are for the most part small. The positive correlations are adding (by amount) 50, (by %) 16, multiplying (amount) 30, (rank) 50, tapping (second method) 06, (Investigation B) 10, crossing test (amount) 27, (by per cent) 20. Color-naming measured by all methods in both experiments, tapping by both methods in Investigation A, word-building, multiplying by the second method, opposites, calculation, discrimination, and the three hole test show negative correlations throughout.

The preponderance of negative over positive correlations in number and amount suggests that, if our measures are accurate ones, greater relative ability at the start, in some of these tests, need not go with greater ability to improve.

A comparison of measures of gain is instructive. Table 28 gives the average correlation of each measure of improvement with the two or three other scores used. The measure called "when reached" has because it is not comparable with the other scores, been omitted here. So improvement in color-naming, measured by the time method has an average correlation of .93 with improvement as measured by the amount, per cent, and rank method. Next to each entry in the table is a figure in parentheses (one, two, three, or four) which represents the rank of the average correlation as compared with other average correlations in the same test by different methods. The (4) next to 93 (time) in color-naming means that in this test (medians being used as measures) the average correlation of time with the other three measures was found to be less high than the average correlation of any one of the other measures with all three. The (1) after per cent shows that per cent had the highest average correlation. Amount and ranks came next.

In this table, as we would expect, all correlations are positive. They are above fifty per cent except in two cases, above sixty except in five cases. Rankings by different measures of gain seem, then, to resemble each other.

The measures have a certain individuality. In all the experiments together per cent has the highest correlation with

other measures of gain—then come in order, ranks, amount, and time. The ranking by per cents is the same—it should be noted—whether the percent is obtained from a measure of amount of time taken or amount of work performed. If one wishes an order of gain which will be the same whichever way it is obtained and which will indicate most closely the results given by other measures, per cent seems to be the one to use. On the other hand, as has been pointed out many times, this figure has the disadvantage of being further removed from the original results than are figures of actual amount of work done or time required.

TABLE NO. 28

Average Correlation of Measures of Gain.

Test	Time	Amount	%	Rank
I. Investigation A—Median used as measure.				
Color-naming	93 (4)	95 (2)	97 (1)	93 (3)
Adding	87 (2)	81 (1)	91 (1)	84 (3)
Tapping	95 (3)	95 (2)	96 (1)	88 (4)
Multiplying	48 (4)	62 (3)	70 (2)	75 (1)
Word-building	92 (2)	92 (3)	94 (1)	91 (4)
II. Investigation A—1st and Best used as measure.				
Color-naming	87 (2)		91 (1)	84 (3)
Adding	89 (2)		94 (1)	87 (3)
Tapping		92 (2)	94 (1)	87 (3)
Multiplying	69 (3)		84 (1)	76 (2)
III. Investigation B—Median used as measure.				
Color-naming	87 (3)	84 (4)	91 (2)	95 (1)
Opposites	70 (3)	67 (4)	78 (2)	79 (1)
Calculations	81 (4)	87 (3)	93 (1)	92 (2)
Tapping	93 (1)	92 (2)	91 (3)	83 (4)
Discrimination	93 (2)	93 (2)	93 (2)	78 (4)
Crossing	80 (2)	88 (1)	79 (3)	78 (4)
Three Hole	66 (4)	72 (3)	83 (1)	75 (2)

The difference between measures of gain can be demonstrated further by a consideration of Table 27 which gives the correlation of different measures of gain with initial ability. This correlation is, except in three cases where they are equal, less strongly negative in the case of per cent than of time. Similarly correlations by amount are more positive than correlations by per cent. The reason for this is inherent in the measures themselves.

A hypothetical case will demonstrate this. Let us suppose that A and B took 25 and 50 seconds respectively to perform a certain task. A, by the time method is better at the beginning than B. Let each one make a gain of five seconds—by the time method an equal gain. If we change these gains into per cents, the equality is lost. A, who was best at first has gained

20%—B who was less skillful has gained only 10%. The initially better person by the per cent method, is given the advantage. So the tendency toward negative correlation, between initial ability and improvement, so marked in the case of the time measure, is reduced by a change to the per cent method, which, when gains equal by the time method are compared allows the person with the better initial record to make the larger gain. The small size of the negative correlation measured by the amount method can be demonstrated in the same way. Turning the original 25 and 50 into terms of amount accomplished—we get 04 and 02. A and B by the original hypothesis each required five seconds less at the end of the practice period. Their times were then 20 and 45 seconds. These turned into amounts give 05 and 022. The one who was better at the beginning is—as in the case of per cent—given the advantage.

But this advantage is greater than the advantage given by the per cent method. In the example the gain of the better person when measured by the per cent method is twice the gain of the poorer, when measured by the amount method, however, this relation is five to two. Amount then will have the least high negative correlation as shown by the fact that the better person, by this method, has the greatest advantage over the poorer. Gains made by persons good at the beginning are raised by a change to the amount method, the negative relation between initial ability and improvability consequently reduced.

Objections are easily raised against the measure which has been called "Rank." Its obvious fault is that it allows for no improvement for the person who is best at first and no loss for the person who is initially worst. Furthermore, rise in position from sixth to third place, for example, is considered equal to a change in rank from 22nd to 19th. Another difficulty was noted when the results were correlated. Almost all gains in rank were small and there were frequently three or four or even five which were equal in amount. Correlations by the rank method are not as accurate when a number of the quantities in one of the orders are equal as when the differentiation of positions is obvious. The rank method when tried out, however, showed a surprisingly high correlation with other measures and a low negative correlation, as did other methods, with initial ability. The evidence from the experiment seems

to show that if the other three scores are valid, that this also gives fairly truthful result. In the present experiment it happened infrequently that the person who was first, remained first, or the person who was last, remained last, so the question of undistributed extremes did not affect the results materially. The assumption that gains at the ends of the scale are equal is one which underlies the other measures as well and so will not affect the correlations between the methods of scoring improvement. It may be argued in favor of this way of measuring, that in daily life gain is more frequently measured in terms of relative than of absolute achievement. The rank method makes improvement contingent on the achievement of others as well as upon the subject's own absolute gain.

The measure called "when reached" is not a measure of improvement in the sense that the others are records of progress. It correlates negatively with other measures of gain, except in the case of adding, a correlation of +02, and in the case of multiplying where there is a small positive correlation throughout. With one exception (tapping) the correlation between initial ability and "when" is low and negative. There seems to be no relation between reaching one's maximum early and being proficient at the start, and a tendency toward negative relation between achieving one's best record early and making a large amount of gain. This latter is a correlation which seems on the face of it, to be a probable one. but, as shown by the small size of the coefficients, not inevitable.

In observing the results gained by the different methods one notes that in three cases out of four the correlation between initial ability and improvement as measured by difference between best scores and initial scores is more strongly negative than the correlation between initial ability and improvement measured by the difference between the medians of the first and last trials. The exception noted is in the case of tapping in whch the two coefficients are approximately equal. Best records and initial records would seem from their derivation to be less reliable than median records. But chance errors would operate to reduce coefficients toward zero rather than toward minus 100. A constant error might be operative here. If it were true that best records always tend to raise a score above its true amount (or first records to lower it unduly), then since there is greater opportunity in the case of the initially poor than of the initially good individuals, such

an error would make the change made by those poorest at the start seem greater than the change made by the initially better people. This error, on the other hand, is only the error to which all unweighted measurements of improvement are subject. We are not discussing the question of opportunity, difficulty, probability, but of actual observable amounts of gain. It is improbable also that best records were frequently much accelerated. Where such increase was due to an "accident in the nervous system" it cannot be called an error, for we are inquiring here not into average performance but ultimate capacity, however caused. The lowering of first records is likewise, if present, an actual fact which should be included in the measurement. The fact that best records have a high intercorrelation and a low negative correlation with improvement, (as do other measures) would seem to argue for their reliability. A possible explanation of the higher negative correlation of best scores might be the fact that the first trial —best trial interval allows for the greatest amount of improvement whereas the median method has, as seen above, irrelevantly admitted opportunity for learning.

Besides the relation between first records and improvement and the relation between the intervals and scores which measure gain, the differences in the tests themselves may be studied—when considered from the point of view of the relation between gain and original ability. We can arrange the tests in a hierarchy using as the basis for our ranking the amount of negative correlation between initial ability and improvability. We may use the "time" method as the measure of gain—as this was the one in which all but three of the original measurements were made. Then we have Table No. 29. In Table No. 30 the arrangement is made by an averaging of the correlations by time, per cent, and amount (where amount was calculated).

That something is actually measured by this hierarchy might be argued from the fact that the same tests by different measures or in different experiments so frequently fall relatively near together in these tables. We have color-naming in Table 29 consistently high –73, –57, –31. Tapping in the same table is always low –15, –03 and 06. Near the tapping test is the crossing tests (–02) which seems to be somewhat similar.

Multiplying has a high negative correlation when measured by either method. Near it are opposites, calculation, and add-

TABLE NO. 29

Negative Correlation of First Records with Improvement as Measured by Time.

Investigation		
B	Opposites	-98
A 2nd Method	Multiplying	-94
B	Calculation	-92
B	Three Hole	-89
A 2nd Method	Adding	-85
A 1st Method	Multiplying	-84
B	Color-naming	-73
A 2nd Method	Color-naming	-57
B	Discrimination	-53
A 1st Method	Color-naming	-31
A 1st Method	Word-building	-28
A 1st Method	Tapping	-15
A 1st Method	Adding	-11
B	Tapping	-03
B	Crossing	-02
A 2nd Method	Tapping	06

TABLE NO. 30

Hierarchy of Negative Correlation of First Records—Measures Averaged

Investigation	Average
A Method 2 Multiplying	-80
A Method 2 Adding	-77
B Opposites	-72
B Calculations	-70
B Three Hole	-54
B Discrimination	-53
B Color-naming	-53
A Method 2 Color-naming	-49
A Method 1 Multiplying	-24
A Method 1 Color-naming	-19
A Method 1 Word-building	-14
A Method 1 Tapping	-13
A Method 2 Tapping	-06
B Tapping	01
B Crossing Test	15
A Method 1 Adding	18

ing (by one method). Adding and calculation differ of course only in the particular form of the operation involved. The similarity between the opposites and the calculation tests has frequently been commented upon. Table No. 30 shows (with some exceptions) a like result. Tapping is consistently low—and ranks with crossing—color-naming is consistently mediocre, multiplying (by one method), opposites and calculation are all high. The discrimination reaction and the three hole test rank with these.

A general tendency is brought out by the two hierarchies. Tests of the higher, "more intellectual" functions seem to gravitate toward the top of the list—those of the lower

"motor" functions are more frequently found below the middle. Opposites, calculation, and multiplying are definitely at the top. Tapping and crossing are always at the bottom, and color-naming is always mediocre. Were the order strictly that of intellectual to less intellectual functions probably the position of word-building would be raised and that of discrimination lowered.

The tendency is sufficiently well marked for us to observe that in tests employing higher mental functions, initial proficiency is less likely to imply subsequent improvement than is ability at the start in lower forms of achievement likely to indicate probable gain to be made in those particular tests.

It is difficult to explain this hierarchy. Closeness of approach to a limit, or relative unreliability of the more complex performances are possible though not probable explanations. Tests at the top of the hierarchy show no consistent tendency to be those in which the end of improvement is most closely approximated. Opposites and calculations, it is true, showed a level of improvement but tapping (Investigation B) did also. Color-naming which showed a definite level stands, not as we might expect, at the top but in the middle. Multiplying —the test in which there was the greatest possibility of gain of all—is not at the bottom but near the top of the list.

Nor does the hierarchy follow the order of reliability. The reliability of the tests in Investigation B is not determined but because of the large amount of practice given it is safe to assume that individual variability has been greatly reduced. Calculations of Investigation B should then be more reliable than adding of Investigation A. It is equally or more strongly negative. Similarly color-naming in Investigation B, with a larger practice period is more negative than color-naming in investigation A. The color-naming and adding tests of the present study contrasted with the more variable multiplying, tapping and word-building tests. Multiplying which is most unreliable is at the top of the list, tapping also unreliable is at the bottom. If multiplying were excluded one might venture the suggestion that the least reliable tests had their coefficients reduced toward zero, whereas in those that are less variable, in which practice is continued for the longest time—the correlations are more strongly negative. All arguments from such a hierarchy are dangerous, not only because of the number of factors involved, but because of the chance errors

which may enter in, the tendency to argue from striking cases, etc. It is sufficient to observe that this hierarchy seems to give results opposite to those obtained by other investigators—showing that improvement in the more complex functions has a larger negative relation to original skill than has improvement in the simpler functions.*

Since the matter of unreliability may still be urged as an objection it might be well to determine improvement by using two of the least variable measures even though this determination reduces enormously opportunity for gain. The practice series (of Investigation A) for this purpose, was divided into halves. Eleven trials were included in each half. Improvement is measured from the median of the first to the median of the second half. Table 31 shows the correlation between the first and second halves, and also the correlation between initial ability, as measured by the median of the first eleven trials, and improvement, as measured by the difference between the median of the first and second halves.

TABLE NO. 31

	Cor. Initial Ability and Improvement	Cor. between two Halves
Color-Naming	–21	.96
Adding	–60	.94
Tapping	34	94
Mul.	–54	87
Word-Building	–40	88

The correlations here are still, with the exception of tapping, negative. They are less in the case of multiplying and color-naming, very much greater in the case of adding. The reduction of the coefficients toward zero is explicable as a result of reducing the amount of practice. A rise in correlation toward minus one hundred might be due to increased reliability of results.

The very high correlation between the two halves of the test is important. In spite of the fact that the poorest individuals are gaining the most—whether gain is measured in terms of actual amount or time gained—they do not, during the interval between the median of the first and of the last

*If the hierarchy were in terms of amount done the order would be different. But as this is the method furthest removed from the original measures such a hierarchy would seem to be of lesser value. What the results would be had the original scores been in terms of amount done and the practice equated in terms of time spent we do not know.

half of the experiment, catch up with those who were initially most proficient. Changes in rank are more noticeable (see Chapter 2) when the median of three trials or one trial only is considered.

In order to make this study comparable to others on the same subject, Table 32 is added, which gives the average gain for the six who were initially most proficient, the five who were least skillful, and the two intervening sets of six. Improvement is measured by the time method over the interval, median of the first to median of the last three trials.

TABLE NO. 32

Indiv.	C. N.	Tap.	Add.	Mul.	W. B.
1-6	11.63	31.1	22.2	101.8	4.65
7-12	12.16	64.5	21.3	403.8	7.0
13-18	14.63	33.8	18.5	623	8.1
19-23	36.69	103.6	27.8	738.6	6.5

Averages point to the same result as did the correlations. The initially better gain the least. Those variations which are observed in this table might easily be chance results due to the small number of cases in each group. If the first two and the last two groups are averaged there is a perfect correspondence between original proficiency and little gain.

The correlations and averages, then, seem to show a general tendency toward negative relationship between initial ability and improvement through practice. This is true whether we measure the gain over long or short intervals of practice, whether in general, we use the time, amount, per cent, or rank method, whether the function studied is the complex multiplying or the simpler color-naming. In general it seems true that the longest intervals and in some cases the apparently more reliable measures give the highest negative correlation. The time method is always most strongly negative, the per cent next, and amount last. The more complex functions show this tendency more markedly than do the simpler traits.

The limits of diagnostic value of the initial or of the first two or three trials of a test are immediately obvious. Such tests are indicative (see Chapter 3) of final status. They are of little value in determining the actual amount or per cent of improvement an individual will make during a given amount of practice. An initial high score should make one predict relatively little rather than relatively great future gain.

CHAPTER FIVE

THE RELATION OF IMPROVEMENT TO FINAL ABILITY.

Closely related to the problem of initial achievement versus probability of improvement is the question of the correlation of final ability and amount of gain. The problem here is—what is the cause of the final high ranking of some persons, and the low position of others? Have those who are eventually best achieved their score through improvement in the practice series and have those who are poor shown a notable lack of gain—or must some other factor be suggested to explain final position?

The answer to this question can be found most easily by correlating the final order with the order of gains. As in the previous study, two intervals were used upon which to reckon gain. Two scores for each of these intervals, the original gross improvement and the percentage of improvement were correlated with final ability. "When reached" the measure of the time at which the maximum record was made, was also correlated with final achievement. Table No. 33 presents the results of this correlation.

TABLE NO. 33

Final Position and Amount of Improvement.

	By Time	%	When
Color Naming			
Final Position & Gain	18	27	
Best Score & Gain	–03	23	–41
Tapping			
Final Position & Gain	64	61	06
Best Score & Gain	70	62	
Multiplying			
Final Position & Gain	–07	.66	
Best Score & Gain	–34	23	11
Adding			
Final Position & Gain	27	51	
Best Score & Gain	08	34	–36
Word-Building			
Final Pos. & Gain	48	39	
Experiment B.			
Color-naming	–03	18	
Opposites	–02	26	
Calculation	21	62	
Tapping	87	91	
Discrimination	18	18	
Crossing Test	72	86	
Three Hole	–77	–26	

In this table there is a preponderance of positive over negative coefficients. Of the 32 figures given there, excluding "When reached," only seven are negative. Except in the case of the three hole test all correlations, when gain is measured by the per cent method, are positive. Except in four cases, the determination by per cent has a higher correlation than the determination by original measures. In one of these cases both have the same correlation, in three the original record was made in terms of amount. It can be demonstrated by the means applied in Chapter 4 for a similar purpose, that the best person at the end is given a greater advantage in reckoning improvement by the amount than by the per cent method, and by the per cent than by the time method. So we should expect correlations by amount to be higher than by per cent which are similarly higher than by time.

This table, then, shows a general tendency toward a positive relation. Ability to improve by practice seems to be a factor in the gaining of a final high position. This ability to improve is best defined, however, by considering it as the capacity to make a gain which is a large per cent of one's original score, or to perform a greater amount of work in a given time rather than as ability to perform a given amount of work very much more quickly. The relative size of the correlation by per cent emphasizes the importance of considering gain—when we speak of its correlation with final position—as an improvement large in relation to one's original score. The finally best person in multiplying does not make a gain of a greater number of seconds than the worst person. Because of the wide range of scores this would be in some cases impossible. But he does gain a large per cent of his original score.

It is important to note that the best score in three cases out of four correlates less closely with improvement than the last score. The exception noted is tapping, which was originally measured by the amount method. The two coefficients are in this case approximately equal. The three other cases seem to show a tendency for the person who is best at the end to owe her position largely to practice gains—while the person who attains a high ranking somewhere in the practice series seems to derive this ability not so surely from improvement made between first and best score. This is, possibly what one would expect. He who is best at the end has had a longer time in which to make large practice gains. The person who is most

efficient when efficiency is measured by a high score made anywhere in the practice period may have made his best record very early in the game and may, with less opportunity for improvement by practice, be depending more upon his native ability. Such a finding is similar to that observed in Chapter 4, where the correlation of gain, measured by initial score—best record interval, also gave a higher negative correlation with initial records. The negative correlation of reaching one's best score early in the series, found in color-naming and adding contrasts with the lack of relation observed in tapping and multiplying. The finally best person in color-naming and adding seems to continue improving for a longer time than other less skillful individuals. In tapping and multiplying there seems to be no correlation between final achievement and time of reaching one's improvement level.

In order to check the results obtained by these two methods, the correlation was obtained between final position, as measured by a very reliable score—the median of the last eleven trials, with gain as measured by the difference between the median of the first and of the last eleven trials, in Experiment "A". Table 31 (see Chapter 4) is repeated here—with the additional middle column.

TABLE NO. 34

	Gain & Median 1st Eleven Trials	Gain & Median Last Eleven Trials	First & Last Eleven Trials
C. N.	–21	–03	96
Tap.	34	54	94
Add.	–60	–25	94
Mul.	–54	–16	87
W. B.	–40	04	88

The correlation between gain and final position is less strongly negative than the correlation between improvement and initial rank. This result is partly due to the mathematics of measuring gain. If we call the original score, a, the gain, g, then final score will equal a+g. Because of the common element g, a+g will be more likely to have a positive correlation with g, than will some other figure.

Improvement measured over this interval shows a tendency, with the exception of tapping, toward a negative or at least toward zero correlation—whereas gain measured over the two longer intervals seemed to be related positively to final score. From such a comparison we must, of course, omit the

scores in terms of per cent and consider only those gains based directly on original records. Though the correlation of gain by the first two methods is not always indisputably positive, it is usually nearer the positive end of the scale than is the correlation by the third method. The order of positiveness is,—first, correlation with gain measured over the interval, median of the first three to median of the last three records, then correlation with gain measured over interval initial score to very best score, and finally gain measured over the interval median of the first eleven to median of the last eleven trials. These intervals differ in reliability and in length. We may say that over the longer, less reliable intervals the relation as measured is frequently positive, over the shorter intervals with more precisely determined score, the relation is more frequently a negative or zero one.

The exception in the case of tapping is notable. In both investigations "A" and "B" final position is seen to be correlated positively with practice gain. One cause for this correlation may be the fact that tapping was measured by the amount method. Or there may be some reason inherent in the test itself. A hierachy of positive correlations (based on original scores measured over the first and second interval) may show something about the difference between the tests.

TABLE NO. 35

Hierachy—Gain correlated with final position—Time used as Measure.

Tapping (B)	.87
Crossing	.72
Tapping (A) (2nd. method)	.70
Tapping (A) (1st Method)	.64
Word-building (1st Method)	.48
Adding (A) (1st Method)	.27
Calculation	.21
Color-naming (A) (1st Method)	.18
Discrimination	.18
Adding (A) (2nd Method)	.08
Opposites	−.02
Color-naming (A) (2nd Method)	−.03
Color-naming (B)	−.03
Multiplying (A) (1st Method)	−.07
Multiplying (A) 2nd Method)	−34
Three Hole	−77

The significance of the hierachy is not very plain. It shows a certain tendency toward an opposite ranking to that of the arrangement in Chapter 5. Tapping—at the bottom before—is at the top here. Opposites is near the bottom instead of the

top. Color-naming which was in the middle, nearer the top in the former arrangement is now nearer the bottom. Discrimination has changed its position markedly. Word-building, multiplying and calculation remain about the same. With the exception of the three hole test and of word-building we might say that the correlation, on the whole, grows less as we approach the more complex, more strictly intellectual functions. These might also be called the less reliable functions.

It is difficult to draw a conclusion from such diverse results. One may emphasize a few observed facts. These are (1) a definite positive correlation between final score and gain when per cent or amount measures are used, (2) a tendency toward postive relationship even when time measures are used if the intervals are sufficiently long, (3) a reduction of this relation toward zero or toward negative correlation when very accurate measures are used at either end of a short practice period, (4) a tendency for the simpler functions to show a stronger positive correlation. One might expect, then, that long continued practice at some relatively simple task, would give, especially if gain were measured in terms of per cent of original score or difference in amount of work performed, a high positive correlation between final position and improvement. More difficult performances, especially if repeated for a short time only and measured in terms of time required might show a small positive relation, no relation or a negative correlation between gain and rank at the end of the series of repetitions.

Of greatest importance in all these discussions are the figures in Tables 3, 4 and 5, (Chapter 2) and Table No. 34 (Chapter 5). Though the relation of final position and gain may vary with accuracy, length of practice, kind of functions studied, the correlation between initial and final position is always a positive one.

CHAPTER SIX

THE RELATION OF IMPROVEMENT IN ONE TEST TO IMPROVEMENT IN ANOTHER AND OF INITIAL AND FINAL GAIN.

The correlation between scores in the different tests and between initial and final position has been considered. Problems of relation may be studied not merely by means of a comparison of individual records but through the correlation of individual improvement. We may ask, does gain in one test imply improvement in one or all of the other tests? Is gain at the beginning of the practice series prognostic of gain at the end?

In order to make the measure of improvement as reliable as possible the amount of practice has been reduced, in obtaining an answer to our first problem, and that measure of gain employed which is the difference between the medians of the first and of the last eleven trials. No correction for attenuation has been made. Correlations of improvement in every test with improvement in every other test give the followng table:

TABLE NO. 36

Correlation of Improvement in Every Test with Improvement in Every Other Test.

	C. N.	Tap.	Add.	Mul.	W. B.
C. N.	x	-.08	-07	-04	18
Tap.	-08	x	12	-21	-18
Add.	-07	12	x	-09	15
Mul.	-04	-21	-09	x	-04
W. B.	18	-18	15	-04	x
Average	-003	-09	03	-10	03

The relation according to this measure of gain, in these particular tests, seems to be purely a chance one. Even such similar tests as adding and multiplying have a correlation which is very nearly zero. There is no evidence here for any general capacity for improvement.

The correlations are similar to those obtained by Chapman and Wimms. The average inter-correlations between Chapman's tests measured by two improvement scores were, for color-naming, adding and multiplying:

TABLE NO. 37

	C. N	Add.	Mul.
C. N.	x	–11	.00
Add.	–11	x	.05
Mul.	.00	–05	x

Wimms** gets a correlation of .007 between improvability in addition and improvability in multiplication.

As Chapman points out* our practice periods are probably too short. Improvement measured from infancy up would probably give a positive correlaton. But over such periods of time as we have been able to measure gain and with such scores as we can obtain, there is no evidence whatsoever of a positive relation between gain in one test and gain in another. This fact contrasts with the discovery that even first trials of tests frequently show a positive correlation. What a person can do in one test is a better indication of what he can do in another, than is his ability to improve in one capacity a sign of ability to make progress in another.

The problem of the correlation between early and late gain is an important one. It might be that gain during the first part of an intensive practice experiment would be evidence of a capacity for improvement along that particular line, so that the individual able to make a larger improvement than his fellows at the beginning of practice would also make larger relative gains at the end of the series. On the other hand it is possible that large improvement at the beginning would, because of the approximation of a physiological limit, preclude large gains at the end, and we would find a negative correlation between initial and final gain. An examination of the relation between gain measured at different parts of the practice period might throw some light upon the value of any measure of improvement as an index of an individual's ability for further improvement.

Correlations between gain in the first part of the experiment (measured by the difference between the medians of the second and fourth divisions) and gain in the last part of the series (measured by the difference between the medians of the sixth and of the eighth divisions) make up the followng table:

*Chapman—Individual Differences in Ability and Improvement, T. C. Contrib.

**Wimms—The Relative Effects of Fatigue and Practice B. J. P. vol. 2.

TABLE NO. 38

Correlation between Initial and Final Improvement.

Test	Correlation
Color-naming	.12
Tapping	–.23
Adding	.04
Multiplying	–.09
Word-building	–.15
Average	–.06

Test	Early & Late Improvement
Color	.19
Addition	.40
Mental Multiplication	.40

The correlation, again, seems to be negative or insignificant. It is noticeable that the less reliable tests give, in general, a higher negative correlation. Chapman's results for the same tests give a positive coefficient.

The color-naming test has a correlation approximately equal to the one found here.

The main difficulties with the scores which have been used are:

I. Individuals start at different points having had more or less previous practice and gained more or less from that experience. 2. Where a physiological limit is reached or approached correlations, after this point is attained by any individual are affected. 3. Our measures of gain even throughout the period of improvement may not be equitable. 4. Our actual scores are frequently unreliable. The coefficients which we have obtained are caused, we believe, by a combination of these factors. They give no evidence for the fact that ability for continued improvement may be possessed by individuals who begin by improving. Nor do they show that gain at first necessarily excludes gain at the end. The chance relationship may be a compound of the two tendencies.

One fact is evident. Improvement measures are of little value for diagnosis. They do not indicate either ability to improve in related tests or ability to improve over a different period of time in the same test. These measures are decidedly inferior in their reliablty and their amount of relationship with other scores to such measures as initial or final records. The former are measures of difference. They depend not only upon proficiency but upon a lack of ability. For an improvement score to be very large, one of the scores from which it is derived must be a poor one. In this, of course, they resemble

"fatigue" scores* and the problem is also analogous to that met with in determining whether "quick learners" are "long retainers." But actual records and average records depend purely upon proficiency. Whereas learning tests, measures of improvements over short periods of time may give no insight into general or special ability to gain, actual records in one test may be indicative of scores to be obtained in another test and, in this experiment, are diagnostic of future rank in the same test.

*For a short discussion of this see Wells' "A Neglected Measure of Fatigue." A. J. P. 1908.

CHAPTER SEVEN

SUMMARY

The results of this study summarized in what seems to be the order of their importance and definiteness are:—

1. Correlations between functions become more positive as those tests become better samples of the individual's true ultimate capacity.

This is demonstrated by the increase in intercorrelations which continues as long as there is improvement through practice. That this change is not entirely due to decreased variability of the indvidual scores is suggested by—

a. The fact that correction of the coefficients for attenuation does not do away with the increase.

b. The very high inter-correlation of those records which best represent ultimate capacity (very best records made in the entire series) in spite of their presumable "unreliability."

2. Correlations between records made in the same function at different parts of the practice series are always positive and usually high. Even beginning and end records show a positive correlation.

3. Correlations between improvement and initial records are usually negative. The longer the interval allowed for gain and the more complex the functon, the higher becomes the negative correlation.

4. Final position, especially when the per cent or amount measure of gain is used, is frequently positively related to improvement. The simpler functions show this more markedly.

5. Correlations between gains made in different functions are only such as might occur by chance.

6. No relation is found between improvement in one part of the practice period and improvement in another part of the practice period, of the same function.

A few minor results are:—

1. Correlation between alternative measures of gain over the same period in the same function are postive and high.

2. Gain measured by the "time required method" correlates more negatively with initial position than does gain

measured by the per cent method which is similarly more negative than gain measured by amount done.

3. The best of three records is, in general, as reliable a score as is the median of those records.

4. A correlation of adjacent scores in the same test confirms the results of other observers. Individual variability decreases as long as improvement in the functions continues.

The implication of these results seems to be that records made near the beginning of practice are indicate of future proficiency in the same function although initial ability may prove to be not diagnostic of but rather opposed to further improvement. Scores made by an individual in one performance are diagnostic of his relative standing in other performances, particularly if the records used are those obtained after a certain amount of improvement through practice has occurred. The reliability of actual scores is contrasted with the lack of agreement with other measures, of records of gain. Gain in one performance bears no relation to gain in another, nor is gain at the beginning of the practice series indicative of further improvement.

We find here then, another demonstration of the positive correlation existing between desirable traits. Our results may be used to show that individual differences decrease through practice, though various difficulties encountered in measuring improvement and the high correlation observed between initial and final position, may indicate that the true conclusion is the opposite of the one that the statistical procedure seems to demonstrate.

A few suggestions for further investigation of this problem might be offered. The precautions to be emphasized are those which provide for the certainty of change and those which prevent the penalizing of the initially better individual.

1. The functions studied should be sufficiently complex so that there is opportunity for improvement. The tapping and color-naming tests of the present experiment are, though they give reliable results, undesirable for the purpose because of the small amount of improvement observed. Mental multiplication would be a good choice for such a study.

2. The amount of practice given should be greater than given in former investigations. Different results are observed even in the present study when gain is measured over differ-

ent intervals. An extension of the repetition period to still greater lengths might change the conclusions.

3. Gain should preferably be measured in terms of amount of work accomplished in a given time.

4. The group should consist, in so far as possible, of unselected individuals.

ARCHIVES OF PSYCHOLOGY

SUB-STATION 84, NEW YORK CITY

In addition to the numbers of the Archives, the following monographs are to be obtained from us:

Measurements of Twins: EDWARD L. THORNDIKE. 50 cents.
Avenarius and the Standpoint of Pure Experience: WENDELL T BUSH. 75 cents.
The Psychology of Association: FELIX ARNOLD. 50 cents.
The Measurement of Variable Quantities: FRANZ BOAS. 50 cents.
Linguistic Lapses: FREDERIC LYMAN WELLS. $1.00.
The Diurnal Course of Efficiency: HOWARD D. MARSH. 90 cents.
The Time of Perception as a Measure of Differences in Sensations: VIVIAN ALLEN CHARLES HENMON. 60 cents.
Interests in Relation to Intelligence: LOUISE E. POULL. $1.00. Reprinted from *Ungraded.*
The Conditioned Pupillary and Eyelid Reactions: HULSEY CASON. $1.00. Reprinted from the *Journal of Experimental Psychology.*

THE JOURNAL OF PHILOSOPHY

SUB-STATION 84, NEW YORK CITY

Published on alternate Thursdays

$4 PER ANNUM, 26 NUMBERS 20 CENTS PER COPY

Edited by Professors F. J. E. WOODBRIDGE and WENDELL T. BUSH of Columbia University.

There is no similar journal in the field of scientific philosophy. It is identified with no philosophical tradition, and stands preeminently for the correlation of philosophy with the problems and experience of the present.

Among the many psychological articles appearing in this Journal are the following of recent date:

Measures of Intelligence: A. T. POFFENBERGER
The Identity of Instinct and Habit: KNIGHT DUNLAP
Must We Give Up Instincts in Psychology? J. R. GEIGER
The Modification of Instinct: WALTER S. HUNTER
Mr. Russell's Psychology: F. C. S. SCHILLER
Intelligence and Intellect: A. A. ROBACK

ARCHIVES OF PHILOSOPHY

Editorial communications should be addressed to

Professor FREDERICK J. E. WOODBRIDGE, Columbia University, New York City

Order should be sent to ARCHIVES OF PHILOSOPHY
Sub-Station 84, New York City

The numbers are as follows:

The Concept of Control. SAVILLA ALICE ELKUS. 40 cents.
The Will to Believe as a Basis for the Defense of Religious Faith. ETTIE STETTHEIMER. $1.
The Individual: A Metaphysical Inquiry. WILLIAM FORBES COOLEY. $1.00
The Ethical Implications of Bergson's Philosophy. UNA BARNARD SAIT. $1.25.
Religious Values and Intellectual Consistency. EDWARD H. REISNER. 75 cents.
Rosmini's Contribution to Ethical Philosophy. JOHN FAVATA BRUNO. 75 cents.
The Ethics of Euripides. RHYS CARPENTER. 50 cents.
The Logic of Bergson. GEORGE WILLIAM PECKHAM. 75 cents.
The Metaphysics of the Supernatural as illustrated by Descartes. LINA KAHN. $1.00.
Idea and Essence in the Philosophies of Hobbes and Spinoza. ALBERT G. A. BALZ. $1.25.
The Moral and Political Philosophy of John Locke. STERLING TOWER LAMPRECHT. $1.50.
Science and Social Progress. HERBERT WALLACE SCHNEIDER. $1.25.

DIRECTORY OF AMERICAN PSYCHOLOGICAL PERIODICALS

American Journal of Psychology—Ithaca, N. Y.; Morrill Hall.
Subscription $6.50. 600 pages annually. Edited by E. B. Titchener.
Quarterly. General and experimental psychology. Founded 1887.

Pedagogical Seminary—Worcester, Mass.: 950 Main Street.
Subscription $5. 575 pages annually. Edited by G. Stanley Hall.
Quarterly. Pedagogy and educational psychology. Founded 1891.

Psychological Review—Princeton, N. J.: Psychological Review Company.
Subscription $4.25. 480 pages annually.
Bi-monthly. General. Founded 1894. Edited by Howard C. Warren.

Psychological Bulletin. Princeton, N. J.: Psychological Review Company.
Subscription $5. 720 pages annually. Psychological literature.
Monthly. Founded 1904. Edited by Shepherd I. Franz.

Psychological Monographs. Princeton, N. J.: Psychological Review Company.
Subscription $5.50 per vol. 500 pp. Founded 1895. Ed. by James R. Angell.
Published without fixed dates, each issue one or more researches.

Psychological Index. Princeton, N. J.: Psychological Review Company.
Subscription $1.50. 200 pp. Founded 1895. Edited by Madison Bentley.
An annual bibliography of psychological literature.

Journal of Philosophy—New York; Sub-Station 84.
Subscription $4. 728 pages per volume. Founded 1904.
Bi-weekly. Edited by F. J. E. Woodbridge and Wendell T. Bush.

Archives of Psychology. Sub-Station 84, New York: Archives of Psychology.
Subscription $5. 500 pp annually. Founded 1906. Edited by R. S. Woodworth.
Published without fixed dates, each number a single experimental study.

Journal of Abnormal Psychology and Social Psychology—Boston.
Subscription $5. Richard G. Badger. Edited by Morton Prince.
Bi-Monthly. 432 pages annually. Founded 1906. Abnormal and social.

Psychological Clinic. Philadelphia: Psychological Clinic Press.
Subscription $2.50. 288 pages. Ed. by Lightner Witmer. Founded 1907.
Without fixed dates (9 numbers). Orthogenics, psychology, hygiene.

Training School Bulletin. Vineland, N. J.: The Training School.
Subscription $1. 160 pages annually. Edited by E. R. Johnstone. Founded 1904.
Monthly (10 numbers). Psychology and training of defectives.

Journal of Educational Psychology. Baltimore: Warwick & York.
Subscription $4. 576 pages annually. Founded 1910.
Monthly (9 numbers). Managing Editor, Harold O. Rugg.
(Educational Psychology Monographs.
Published separately at varying prices. Same publishers.)

Comparative Psychology Monographs—Baltimore: Williams & Wilkins Co.
Subscription $5. 500 pages per volume. Edited by W. S. Hunter.
Published without fixed dates, each number a single research.

Psychoanalytic Review. Washington, D. C.: 3617 10th Street, N. W.
Subscription $6. 500 pages annually. Psychoanalysis.
Quarterly. Founded 1913. Edited by W. A. White and S. E. Jelliffe.

Journal of Experimental Psychology—Princeton, N. J.
Psychological Review Company. 480 pages annually. Experimental.
Subscription $4.25. Founded 1916. Bi-monthly. Ed. by John B. Watson.

Journal of Applied Psychology. Worcester, Mass.: Florence Chandler.
Subscription $4. 400 pages annually. Founded 1917.
Quarterly. Edited by James P. Porter and William F. Book.

Journal of Comparative Psychology—Baltimore: Williams & Wilkins Co.
Subscription $5. 500 pages per volume. Founded 1921.
Bi-monthly. Edited by Knight Dunlap and Robert M. Yerkes.

This book may be kept.....weeks.

A fine of two cents will be charged for each day books or magazines are kept overtime.

Two books may be borrowed from the Library at one time.

Any book injured or lost shall be paid for by the person to whom it is charged.

No member shall transfer his right to use the Library to any other person.